Computer intensive methods & spatial statistics

Berthelsen, Kasper Klitgaard

Publication date:
2003

Document Version
Publisher's PDF, also known as Version of record

Computer Intensive Methods & Spatial Statistics

with an emphasis on perfect simulation and spatial point processes

Kasper Klitgaard Berthelsen

Preface

This thesis is the result of my Ph.D. studies at the Department of Mathematical Sciences, Aalborg University, Denmark. The thesis deals with computer intensive methods in spatial statistics with an emphasis on perfect simulation and point processes.

Chapter 1 contains an introduction to the basics of simulation in statistics and in particular to perfect simulation using coupling. Further, Chapter 1 contains an introduction to point processes including a number of specific types of point processes and concludes with a section on the simulation of point processes including simulation methods not covered by Chapters 2–6.

Chapters 2–6 contain the following papers and manuscripts writting during my Ph.D. study. The number next to each title refers to the corresponding chapter in the thesis.

2 Berthelsen, K.K. & Møller, J. (2002a). Spatial jump processes and perfect simulation. In *Morphology of Condensed Matter*, Red. K. Mecke & D. Stoyan, Lecture Notes in Physics, Vol. 600, Springer-Verlag, 391–417.

3 Berthelsen, K.K. & Møller, J. (2002b). A primer on perfect simulation for spatial point processes. *Bulletin of the Brazilian Mathematical Society*, **33**, 351–367.

4 Berthelsen, K.K. & Møller, J. (2003a). Likelihood and non-parametric Bayesian MCMC inference for spatial point processes based on perfect simulation and path sampling. *Scandinavian Journal of Statistics*, **30**, 549–564.

5 Berthelsen, K.K., Møller, J., Pettitt, A.N., & Reeves, R. (2003). Markov chain Monte Carlo without the normalising constant. Manuscript (in preparation).

6 Berthelsen, K.K., & Møller, J. (2003b). Non-parametric Bayesian inference for inhomogeneous pairwise interaction point processes. Manuscript (in preparation).

Each of the above texts is self-contained. For published papers the version given here is essentially the one published except for changes to the general layout and section numbering. As a result the notation is only consistent within each chapter.

I developed the necessary simulation software and performed the required simulation experiments for all papers and manuscipts above. The problems treated in papers corresponding to Chapters 2–4 were posed by J. Møller. The formulation of the crucial ideas and the writing of the papers were done together with J. Møller. The paper in Chapter 5 is in preparation. The idea for the

auxiliary variable method was proposed by J. Møller. The application of the auxiliary variable method to point processes was developed in collaboration with J. Møller. I wrote most of the paper in its present form. In its final form the paper will contain a contribution from A.N. Pettitt and R. Reeves (Queensland University of Technology) consisting of an application of the auxiliary variable method to an autologistic distribution. The paper in Chapter 6 is in preparation. The problem was proposed by J. Møller. In its present form I wrote most of the the paper.

My Ph.D. study was carried out under the supervision of Professor Jesper Møller, who has been encouraging all through my study and has always been a great source of inspiration and generous with his time.

As part of my Ph.D. study I spent three months at CWI in Amsterdam, the Netherlands, visiting Dr. Marie-Colette N.M. van Lieshout. I would like to thank her and Radu S. Stoica for many interesting discussions and making the stay a pleasant one.

Aalborg, Denmark, August 2003 Kasper Klitgaard Berthelsen

A few errors have been corrected in this revised version.

Aalborg, Denmark, August 2004 Kasper Klitgaard Berthelsen

Summary

This thesis deals with computer intensive methods in spatial statistics with emphasis on perfect simulation and spatial point processes. Perfect simulation algorithms are a class of algorithms which return a sample from a specific distribution in finite time with probability one. The term "perfect" is preferred over "exact" in order to emphasize the limitations induced by defective random number generators and limited computer precision. Further, perfect simulation may not deliver an output within the practical constraints of time, which in some cases may lead to so-called user-impatience bias. Limitations of computer storage may be another problem when using perfect simulation. The development of new and more complex algorithms for perfect simulation has been greatly promoted by the seminal work by Propp and Wilson (1996). So far, perfect simulation has been particularly successful within spatial statistics and not least for spatial point processes. Spatial point processes typically serve as stochastic models for point patterns, which typically represent the locations of objects in two or three dimensional space.

An introduction is given to general simulation techniques with an emphasis on techniques based on Markov chains including the Metropolis-Hastings algorithm and Propp-Wilson's coupling from the past algorithm for perfect simulation. Furthermore, an introduction to point processes and to the simulation of these are given. Locally stable pairwise interaction point processes are the main focus in this thesis. We demonstrate that it is infeasible to simulate these point processes using rejection sampling and review Wilson's read-once algorithm as an alternative to Kendall and Møller's dominated coupling from the past algorithm. The latter is the primary algorithm for perfect simulation of locally stable point processes in this thesis.

Dominated coupling from the past and a method based on clans of ancestors are compared in terms of their suitability for perfect simulation of locally stable pairwise interaction point processes. It is concluded that dominated coupling from the past is the better choice for this particular simulation task.

A range of spatial jump processes are studied with a focus on birth-and-death and birth-and-catastrophe processes. Schemes for simulation of spatial jump processes by a coupling to a simple jump processes on the non-negative integers are considered. Point processes specified by a density w.r.t. a "geometric point process" are introduced. For this class of point processes a particular birth-and-catastrophe process proves useful as the basis of a simple perfect simulation algorithm which is in fact of the read-once type.

A combination of path sampling and perfect simulation of point processes for estimating ratios of unknown normalising constants is studied. This approach is used for estimating the unknown normalising constant for a particular parametric spatial point process for a grid of parameter values. This in turn provides a tool for likelihood inference. Further, a non-parametric Bayesian analysis of

a data set is carried out, where we assume that the data can be modelled by a pairwise interaction point process. A priori the first order term is assumed constant and the second order term is assumed to be a random increasing step function. The resulting posterior mean is used as a smooth estimate of the second order term. Simulation of the posterior is done using a Metropolis-Hastings algorithm involving ratios of unknown normalising constants. Here perfect simulation proves infeasible when ratios of unknown normalising constants in the Hastings ratio are estimated. Instead a non-perfect approach is applied. The quality of the non-perfect approach is studied by comparing results obtained using the non-perfect approach with results obtained using the method based on perfect simulation for a simulated data set.

Using a Metropolis-Hastings algorithm for posterior simulation often involves a ratio of unknown normalising constants in the Hastings ratio. An auxiliary variable method is introduced which avoids approximating this ratio of unknown normalising constants. The method is simple but requires perfect simulation of the likelihood. For the algorithm to be successful the auxiliary variable has to be a good approximation of the likelihood. A number of simulation studies are carried out to clarify the advantages and limitations of the method. In a point process setting we find that a particular partially ordered Markov model works well as the auxiliary variable for pairwise interaction point processes.

The last part of the thesis deals with an analysis of an inhomogeneous point pattern in a Bayesian framework. A priori it is assumed that data can be modelled be a pairwise interaction point process with an inhomogeneous first order term of a location-independent second order term. The first order term is a priori assumed to be given by a shot noise process. The method is applied to a data set where simulation of the posterior is done using the auxiliary variable method above.

Sammendrag

Denne afhandling omhandler computer intensive metoder i rumlig statistik med vægt på perfekt simulation og rumlige punktprocesser. Perfekt simulation er en klasse af algoritmer der returnerer en stikprøve fra en given fordeling i endelig tid med sandsynlighed en. Betegnelsen perfekt er foretrukket fremfor eksakt for at understrege begrænsninger forårsaget af defekte tilfældighedsgeneratore og begrænset computerpræcision. Desuden kan perfekt simulation i nogle situationer vise sig at være praktisk uanvendeligt pga. for lange beregningstider og begrænset computerlagerplads. Udviklingen af nye og mere komplekse algoritmer til perfekt simulation er i høj grad blevet fremmet af det banebrydende arbejde af Propp og Wilson (1996). Til dato har perfekt simulation særligt haft succes i rumlig statistik og ikke mindst for rumlige punktprocesser. Rumlige punktprocesser tjener typisk som stokastiske modeller for punktmønstre, der typisk repræsenterer placeringen af objekter i det to- eller tredimensionale rum.

Der gives en introduktion til generelle simulationsteknikker med vægt på metoder baseret på Markov kæder herunder Metropolis-Hastings algoritmen og Propp-Wilsons coupling from the past (CFTP) algoritme til perfekt simulation. Desuden gives en introduktion til punktprocesser og simulationen af disse. I denne afhandling er fokus primært rettet mod lokalt stabile parvis interaktionspunktprocesser. Vi demonstrerer at rejection sampling i praksis er uanvendelig til simulation af punktprocesser og betragter desuden Wilsons read-once algoritme som et alternativt til Kendall og Møllers domineret CFTP. Sidstnævnte algoritme er den primære algoritme til perfekte simulation af lokalt stabile punktprocesser i denne afhandling.

Til perfekt simulation af lokalt stabile parvis interaktionspunktprocesser sammenlignes domineret CFTP og en metode baseret på såkaldte clans of ancestors. Det konkluderes at domineret CFTP er det bedste valg til denne simulationsopgave.

En række rumlige springprocesser studeres med fokus på fødselsdøds- og fødselskatastrofeprocesser. Metoder til simulation af rumlige springprocesser vha. en kobling til simple springprocesser på de ikke-negative heltal betragtes. Punktprocesser specificeret ved en tæthed mht. en "geometrisk punktproces" introduceres. For denne klasse af punktprocesser viser en særlig fødselskatastrofeproces sig brugbar som udgangspunkt for en perfekt simulationsalgoritme af den såkaldte read-once type.

En metode til estimation af en kvotient mellem to ukendte normeringskonstanter baseret på en kombination af path sampling og perfekt simulation af punktprocesser studeres. Denne metode anvendes til estimation af ukendte normeringskonstanter for en særlig parametrisk rumlig punktproces for et gitter af parameterværdier. Dette anvendes efterfølgende til likelihood-inferens. Desuden udføres en ikke-parametrisk analyse af et datasæt, hvor det antages at data kan modelleres af en parvis interaktionspunktproces. A priori antages

førsteordensledet konstant og andenordensledet antages at være en voksende trin-funktion. Den resulterende a posteriori middelværdi anvendes som et glat estimat af andenordensledet. Til Simulation af a posteriori fordelingen anvendes en Metropolis-Hastings algoritme, der involverer en kvotient mellem to ukendte normeringskonstanter. Her viser perfekt simulation sig at være for beregningskrævende i forbindelse med ovennævnte estimationsmetode. I stedet anvendes en ikke-perfekt metode. Kvaliteten af den ikke-perfekte metode studeres ved at sammenligne resultater opnået ved anvendelsen af den ikke-perfekte metode med resultater opnået ved anvendelse af metoden baseret på perfekt simulation for et simuleret datasæt.

Ved anvendelse af en Metropolis-Hastings algoritme til simulation af a posteriori fordelingen forekommer det ofte, som i tilfældet ovenfor, at dette involverer en kvotient mellem to ukendte normeringskonstanter. En metode introduceres, der undgår estimationen af kvotienten mellem to ukendte normeringskonstanter ved at anvende en såkaldt hjælpevariabel. Metoden er simpel, men kræver perfekt simulation af likelihood-ledet og for at være praktisk anvendelig kræves det, at hjælpevariablen er en god tilnærmelse af likelihood-ledet. Et antal simulationsstudier udføres for at afdække fordele og ulemper ved denne metode. For punktprocesser viser det sig, at en særlig partially ordered Markov model fungerer godt som tilnærmelse af en parvis interaktionspunktproces.

Den sidste del af afhandlingen omhandler en Bayesiansk analyse af et inhomogent punktmønster. A priori antages det at data kan modelleres af en parvis interaktions punktproces med et inhomogent førsteordensled og et steduafhængigt andenordensled. Førsteordensledet antages a priori at være givet ved en såkaldt shot-noise proces. Denne tilgangsvinkel anvendes på et datasæt, hvor ovennævnte metode anvendes til simulation af a posteriori fordelingen.

Contents

Introduction

Computers are a vital tool for modern statistics. In some cases they perform simple tasks to avoid otherwise tedious manual calculations. In other cases statistical problems lead to quite computationally intensive methods demanding access to large amounts of computing power to be feasible. This thesis deals with the latter case where the computer intensive aspect is mainly related to simulation based methods for spatial point processes. For these methods often most resources in terms of time and storage are used for the simulation of the point processes and to a lesser degree on using the results of the simulations. For very complicated models, the evaluation of a density may require some computational effort but usually little compared to the actual simulation of the model, see e.g. Chapter 4 and Møller & Waagepetersen (2003).

This introductory chapter is organised as follows. Sections 1.1–1.7 give an introduction to simulation and related topics including Markov chains, the Metropolis-Hastings algorithm and the Propp-Wilson perfect simulation algorithm. Sections 1.8–1.9 contain an introduction to spatial point processes including a brief presentation of a number of widely applicable and flexible classes of spatial point processes. Furthermore, we review some techniques for point process simulation including rejection sampling and Wilson's read-once algorithm, which are not covered by Chapters 2–6.

1.1 Simulation and Monte Carlo estimation

Assume for a moment that we have methods for producing i.i.d. samples from various statistical models of interest. The most basic use of simulation is then to conduct an exploratory study of one or more models. For each model a number of realisations are simulated giving a first impression of the nature of the model which may be helpful in an initial model selection phase.

When looking for more detailed insight into a model, it is often only for quite simple models it is possible to e.g. derive closed form expressions for distributions, means and variances of relevant statistics. Such simple models tend to be too simplistic to constitute a satisfactory model for any real world problem. This is the case not least for spatial point processes. For models where insufficient theoretical insight is available we may turn to simulation for answers — this is typically the case for the models considered in this thesis.

Decidning to use a simulation based method the problem is how to do the actual simulation. Even for relatively simple models we might be lack a method for simulating independent realisations or at least a simple, feasible and easily implemented method. In most situation it is possible to simulate an arbitrarily good approximative sample given enough time — the limitation being that it is usually far from clear what "enough time" is.

The most widely applicable algorithm for (approximative) simulation is based on so-called Markov chains and was introduced by Metropolis, Rosenbluth, Rosenbluth, Teller & Teller (1953) and generalised by Hastings (1970), see Section 1.4 for a brief review of the algorithm. Even though this algorithm has been known for decades it is just recently that sufficient computing power has become available to make it feasible to study truly complex statistical models. This development has further lead to an increasing interest in the use of Bayesian statistic which by nature often results in complex models which are intractable from a mathematical point of view. Chapters 4–6 contain examples of this.

In the sequel, we consider the following setup. Let X be a random variable with distribution Π on a state space Ω which for technical reasons is assumed to be equipped with a separable σ-algebra $\mathcal{F}$. We then write $X \sim \Pi$ and for any statistic $g(X)$ of interest let $\mathbb{E}_\Pi g(X)$ denote the expectation of $g(X)$. In some situations we may assume that Π has a probability density $\pi(x) = c^{-1} f(x)$ w.r.t. to some appropriate measure ν and write $X \sim \pi$ or $X \sim f$. Here f is an unnormalised density with normalising constant $c = \int_\Omega f(x)\nu(dx)$. We may write $f(x|\theta)$ and $c(\theta)$ to emphasise the dependence on some model parameter θ. In the following, we refer to the density π, we want to sample from as the target density to distinguish it from other densities involved in a given sampling scheme.

Many simulation based methods basically aim at estimating one or more expectations. Assume that $\mathbb{E}_\Pi g(x)$ exists and we have a method for producing independent identically distributed (i.i.d.) realisations $X_0, X_1, X_2, \ldots$ with common distribution Π. Then

$$\bar{g}_n = \frac{1}{n+1} \sum_{i=0}^{n} g(X_i), \quad n \in \mathbb{N}_0 \equiv \{0, 1, 2, \ldots\} \tag{1.1}$$

is an unbiased estimate of $\mathbb{E}_\Pi g(X)$ refered to as a simple Monte Carlo estimate. This estimate is associated with some uncertainty referred to as the Monte Carlo error. In the i.i.d. case considered above the classical strong law of large numbers tells us that $\lim_{n\to\infty} \bar{g}_n = \mathbb{E}_\Pi g(X)$ and the classical central limit theorem tells us that $\sqrt{n}\bar{g}_n$ converges in distribution to a normal distribution with mean $\mathbb{E}_\Pi g(X)$ and variance σ_g^2 if the unknown true variance $\sigma_g^2 = \mathbb{V}\mathrm{ar} g(X)$ is finite.

In this section we review a number of methods for simulating different distributions. For most basic univariate distributions clever techniques are available for simulating samples using very little computational effort. A number of these methods are presented and implemented in Press, Teukolsky, Vetterling & Flannery (1992). In the remainder, focus is on methods that are of iterative

or recursive nature where the feasibility relies on the access to a large amounts of computing power.

Since the seminal paper by Propp & Wilson (1996) a lot of work has gone into reworking the Markov chain approach to produce exact samples. In Section 1.6, we review a number of basic algorithms for perfect simulation.

1.2 Rejection sampling

A straight forward and widely applicable (at least in principle) method for simulating independent samples from a density is rejection sampling (von Neumann 1951). Assume that the target process is specified by an unnormalised density f w.r.t. some appropriate measure ν. Assume that another unnormalised density h is available with the properties that it is possible to generate samples from h and that h covers f in the sense that there exists a positive constant $k < \infty$ so that $f(x) \leq kh(x)$ for all $x \in \Omega$. Rejection sampling now works as follows:

1. Generate $x \sim h$.
2. With probability $f(x)/kh(x)$ return x otherwise go to step 1.

It is easily shown that x is distributed according to f. Furthermore, the mean probability for acceptance in step 2 is

$$\int_\Omega \int_0^{f(x)/(kh(x))} c^{-1} h(x) \mathrm{d}u \nu(\mathrm{d}x), \tag{1.2}$$

where c is the possibly unknown normalising constant for h. From (1.2) it is clear that h must be a good approximation of f for this to work well. In general, rejection sampling becomes increasingly difficult as the dimensionality of the target distribution increases. This is illustrated in Section 1.9.3 where rejection sampling is applied to spatial point processes.

As the key to a successful rejection sampler is a good covering density Gilks & Wild (1992) consider an adaptive rejection sampler for univariate log-concave densities. In this scheme the proposed samples are used for sampling as well as modifying the covering density such that it gets increasingly close to the target density.

1.3 Markov chains

For moderately complex models, it is typically either not possible or just not feasible to generate independent samples from the model. In such cases the only solution may be methods based on *Markov chains*. In words, a Markov chain is a sequence (or chain) of random variables $\{X_i : i \in \mathbb{N}_0\}$ where the distribution of the present (or current) state X_i, say, given the past $X_0, \ldots, X_{i-1}$ only depends on the immediate past, i.e. X_{i-1}. Given a distribution Π the aim is to construct a Markov chain in such a way that as i tends to infinity X_i converges

in distribution to Π independent of the initial distribution of X_0. We will make this more precise later. Under certain conditions on the Markov chain $\{X_i : i \in \mathbb{N}_0\}$ it can be proven that the *ergodic average* $\bar{g}_n = (1/(n+1)) \sum_{i=0}^{n} g(X_i)$ tends to $\mathbb{E}_\Pi g(X)$ as n tends to infinity. Hence, the Markov chain replaces the i.i.d. sample in (1.1) giving rise to the now familiar term Markov chain Monte Carlo (MCMC). In the following we briefly consider *time homogeneous* Markov chains in more details. Unless otherwise stated we refer to (Meyn & Tweedie 1993) for proofs and a presentation of the theory on Markov chains on general state spaces. Norris (1997) gives a more legible introduction to Markov chains but in a less general setup.

Markov chains can be described in terms of *transition kernels*. $P(\cdot, \cdot)$ is a transition kernel if $P(x, \cdot)$ is a probability measure on $\mathcal{F}$ for all $x \in \Omega$ and $P(\cdot, F) : \Omega \mapsto [0, 1]$ is measurable for all $F \in \mathcal{F}$. Here time homogeneous means that the transition kernel is independent of time, i.e. conditional on X_i the distribution of X_{i+1} is independent of i.

Given a transition kernel and an initial distribution Π_0 the distribution $\mathcal{D}$ of the Markov chain $\{X_i : i \in \mathbb{N}_0\}$ on the space $(\Omega^\infty, \mathcal{F}^\infty)$ is given by $\mathcal{D}(X_0) = \Pi_0(X_0)$ and $\mathcal{D}(X_i | X_0 = x_0, \ldots, X_{i-1} = x_{i-1}) = P(x_{i-1}, X_i)$.

For all $x \in \Omega$ and $F \in \mathcal{F}$ let $P^n(x, F) = \mathbb{P}(X_n \in F | X_0 = x)$ denote the n-step transition probability. A distribution Π is then an *equilibrium distribution* if $\lim_{n\to\infty} P^n(x, F) = \Pi(F)$ for all $F \in \mathcal{F}$ and almost all $x \in \Omega$. Given a distribution Π it is often possible to construct a transition kernel so that Π becomes the equilibrium distribution of the resulting Markov chain.

A distribution Π is said to be *invariant* for the transition kernel P if $\Pi(F) = \int_\Omega P(x, F)\Pi(\mathrm{d}x)$ for all $F \in \mathcal{F}$, i.e. $X_i \sim \Pi$ implies $X_{i+1} \sim \Pi$. If Π is an equilibrium distribution then Π is also an invariant distribution for the same transition kernel.

A Markov chain with transition kernel P is *reversible* w.r.t. the distribution Π if $\int_B P(x, A)\Pi(\mathrm{d}x) = \int_A P(x, B)\Pi(\mathrm{d}x)$, i.e. if $X_i \sim \Pi$ then (X_i, X_{i+1}) and (X_{i+1}, X_i) are identically distributed. Hence we can generate the Markov chain for increasing as well as decreasing time using the same transition kernel. Furthermore, reversibility implies invariance and is often easier verified than verifying invariance directly.

Assume that it is possible to divide the state space Ω into $k > 1$ disjoint subsets $F_1, \ldots, F_k$ so that $P(x, F_j) = 1$ for all $x \in F_{j-1}$, $j = 2, \ldots, k$ and $P(x, F_1) = 1$ for all $x \in F_k$ then the Markov chain is said to be *periodic*. If the Markov chain is not periodic it is called *aperiodic*. It is intuitively clear that a periodic Markov chain is not converging to any distribution.

A Markov chain is said to be ζ-irreducible if ζ is a non-zero measure on Ω so that for all $x \in \Omega$ and $F \in \mathcal{F}$ with $\zeta(F) > 0$ we have $P^n(x, F) > 0$ for some $n \in \mathbb{N} \equiv \{1, 2, \ldots\}$. If Π is invariant for a ζ-irreducible Markov chain then the Markov chain is also Π-irreducible and Π is the unique invariant distribution. In words Π-irreducible means that the Markov chain can get from any state in the state space to any region of the state space which is assigned a positive probability.

Recall that when we introduced equilibrium distribution we assumed conver-

gence for almost all initial states. A Markov chain with invariant distribution Π is *Harris recurrent* if for all $x \in \Omega$ and all $F \in \mathcal{F}$ with $\Pi(F) > 0$ we have $\mathbb{P}(X_n \in F \text{ for some } n | X_0 = x) = 1$. Harris recurrence implies a *strong law of large numbers* for Markov chains. Assume that $\{X_i : i \in \mathbb{N}_0\}$ is a Harris recurrent Markov chain with invariant distribution Π and the expectation $\mathbb{E}_\Pi g(X)$ exists. Then with probability one $\lim_{n\to\infty} \bar{g}_n = \mathbb{E}_\Pi g(X)$. Note that this result is independent of the initial distribution and we do not need to assume aperiodicity. If a Markov chain is both Harris recurrent and aperiodic it is called *ergodic*.

For any signed measure ζ on Ω let $\|\nu\|_{TV} = \sup_{F\in\mathcal{F}} |\zeta(F)|$ denote the *total variation norm*. A Markov chain with invariant distribution Π is ergodic if and only if $\lim_{n\to\infty} \|P^n(x,\cdot) - \Pi(\cdot)\|_{TV} = 0$ for all $x \in \Omega$. In other word, an ergodic Markov chain converges for any initial state $x \in \Omega$. This results implies that we may get an arbitrarily good approximative sample from Π if we wait long enough. The question is how long is "long enough". We return to this question in Sections 1.5 and 1.6.

A Harris recurrent Markov chain with invariant distribution Π is *geometrically ergodic* if there exists positive constant $r < 1$ and and some function $G(x) < \infty$ so that $\|P^n(x,\cdot) - \Pi(\cdot)\|_{TV} \leq G(x)r^n$ for all $x \in \Omega$ and $n \in \mathbb{N}$. If $G(x) \equiv G$ is constant the Markov chain is uniformly ergodic.

If the Markov chain $\{X_i : i \in \mathbb{N}_0\}$ is geometrically ergodic with invariant distribution Π we have the following *central limit theorem*. Let $g : \Omega_+ \to \mathbb{R}$. If either **i)** $\mathbb{E}_\Pi |g(X)|^{2+\epsilon} < \infty$ or **ii)** the chain is reversible w.r.t Π and $\mathbb{E}_\Pi g^2(X) < \infty$ then $\sqrt{n}(\bar{g}_n - \mathbb{E}_\Pi g(X))$ converges in distribution to a normal distribution with mean zero and variance $\bar{\sigma}_g^2$. The asymptotic variance $\bar{\sigma}_g^2$ is given by

$$\bar{\sigma}_g^2 = \sigma_g^2[1 + 2\sum_{m=1}^{\infty} \rho_g^{(m)}],$$

where the variance $\sigma_g^2 = \mathbb{V}\text{ar}(g(X_0))$ and the lag m auto-correlation $\rho_g^{(m)} = \mathbb{C}\text{orr}(g(X_0), g(X_m))$ are defined under the assumption that $X_0 \sim \Pi$. For proof of **i)** see Chan & Geyer (1994), for **ii)** see Roberts & Rosenthal (1997).

An unbiased estimate of the asymptotic variance can be obtained using so-called batch means, see Ripley (1987) for more details. Geyer (1992) gives another unbiased estimate for irreducible and reversible Markov chains based on a sequence of estimates for the lag m auto-correlation consistent with theoretical properties for the auto-correlations.

1.4 Metropolis-Hastings Algorithm

Assume that the distribution of interest is specified by an unnormalised density f. A flexible and widely popular algorithm for generating Markov chains with an equilibrium distribution given by f is the Metropolis-Hastings algorithm (Hastings 1970). As with a lot of clever simulation techniques in statistics it originates from statistical physics (Metropolis et al. 1953); Hastings (1970)

extended the method and introduced it in statistics.

The Metropolis-Hastings algorithm generates a Markov chain by an accept-reject mechanism where each update consists in, given the current state, generating a proposal for the next state in the Markov chain. If the proposal is rejected the Markov chain stays in the current state, i.e. the next state is set equal to the current state. This is unlike rejection sampling where each update consists in generating proposals until acceptance.

1.4.1 Fixed dimensional case

Assume that the state space is a d dimensional product space $\Omega = \Omega_1 \times \cdots \times \Omega_k$ where the spaces $\Omega_1, \ldots, \Omega_k$ may be of different dimensions. Then any state $x \in \Omega$ is written as k components $x = (x^{(1)}, \ldots, x^{(k)})$, where $x^{(j)} \in \Omega_j$, $j = 1, \ldots, k$.

Assume that $X_i = x \equiv (x^{(1)}, \ldots, x^{(k)}) \in \Omega$, $i \in \mathbb{N}_0$, is the current state of the Markov chain. A Metropolis-Hastings update then consists in first picking j uniformly randomly from $\{1, \ldots, k\}$. Given j generate a proposal $x'^{(j)}$ for the jth component from a proposal distribution $q_j(x, x'^{(j)})$ which may depend on the current state. Finally, the proposal is accepted with probability $\alpha(x, x')$, where $x' = (x^{(1)}, \ldots, x^{(j-1)}, x'^{(j)}, x^{(j+1)}, \ldots, x^{(k)})$). It is then relatively easily checked that the resulting Markov chain is reversible w.r.t. Π if $\alpha = \min\{1, H(x, x')\}$ where

$$H(x, x') = \frac{f(x')q_j(x', x^{(j)})}{f(x)q_j(x, x'^{(j)})} \tag{1.3}$$

is the *Hastings ratio*. Under additional assumptions the resulting Markov chain is uniformly ergodic. Notice that (1.3) only involves a ratio of densities, hence the Metropolis-Hastings algorithm works even if the normalising constant of f is unknown.

The description above can be summarised in algorithmic form. For $i \in \mathbb{N}_0$ assume $X_i = (x^{(1)}, \ldots, x^{(k)})$. Then

1. Generate $j \sim \text{uniform}\{1, \ldots, k\}$ and given j generate $x'^{(j)} \sim q_j(X_i, x'^{(j)})$.
2. With probability $\alpha = \min\{1, H(x, x')\}$ set $X_{i+1} = x'$ otherwise set $X_{i+1} = X_i$, where $x' = (x^{(1)}, \ldots, x^{(j-1)}, x'^{(j)}, x^{(j+1)}, \ldots, x^{(k)})$.

Instead of generating j uniformly one could use any another distribution on $\{1, \ldots, k\}$. If this distribution depends on the current state of the Markov chain a minor modification of the Hasting ratio is needed. Another alternative is a systematic update scheme where components are updated in some prespecified order. In this case the resulting Markov chain is typically not reversible.

A particular simple version of the algorithm above is the *Metropolis algorithm* introduced by Metropolis et al. (1953). In this case the proposal distributions are assumed symmetric in the sense that $q_j(x, x'^{(j)}) = q_j(x', x^{(j)})$ implying an acceptance probability $\alpha = \min\{1, f(x')/f(x)\}$. As a result proposals of states more likely (according to f) than the current state are always accepted.

Another special case of the Metropolis-Hastings algorithm is the *Gibbs sampler* — a term coined in the paper by Geman & Geman (1984). In statis-

tical physics this version of the Metropolis-Hastings algorithm is known as the heat bath algorithm. For the Gibbs sampler the proposal density for the jth component is the conditional distribution of the jth component given the $k-1$ other components, i.e. $q_j(x, x'^{(j)}) \propto f(x'^{(j)}|x^{(-j)})$ where $x^{(-j)} = (x^{(1)}, \ldots, x^{(j-1)}, x^{(j+1)}, \ldots, x^{(k)})$. We then get $H(x, x') \equiv 1$, i.e. proposal are always accepted. In practice it may be difficult to sample from the conditional distributions. We briefly return to this point in Section 1.9.4.

If the proposals are independent of the current state x the resulting algorithm is known as an *independence sampler*. This proposal type and others not considered here are reviewed in (Dellaportas & Roberts 2003).

1.4.2 Trans dimensional case

In the Metropolis-Hastings algorithm above all transitions were between states of equal dimension. The reversible jump algorithm introduced by Green (1995) gives a general frame work for generating reversible Markov chains allowing transitions between states of different dimension.

Assume for simplicity that we are in the one component setup. Let x be the current state of the Markov chain assumed to be of dimension n. Then with probability $r_m(x)$ we propose a dimension changing move of type m. For most practical purposes we may assume that conditional on m the proposal x' has dimension $n_m > n$ and is given by an invertible deterministic function $\omega(x, u)$ where u is a random variable with density q_m. Note that as ω is invertible u has dimension $n_m - n$. The proposal x' is then accepted with probability $\alpha(x, x') = \min\{1, H(x, x')\}$ where

$$H(x, x') = \frac{f(x')r_{m'}(x')}{f(x)r_m(x)q_m(u)} \left| \frac{\partial \omega(x, u)}{\partial (x, u)} \right|. \tag{1.4}$$

Here m' indicate the type of change of dimension move needed to return to x from x'. Reversibility is then obtained by letting the acceptance probability of the reverse move be $\alpha(x', x) = \min\{1, 1/H(x, x')\}$.

The spatial birth-and-death process given by Geyer & Møller (1994) considered in Section 1.9.5 is essentially a special case of this setup where the dimension changing moves consist in a single point being either added to or removed from a point configuration.

1.5 MCMC in practice

Assume that we have generated an ergodic Markov chain $X_0, \ldots, X_n$ with the purpose of estimating some expectation using Monte Carlo. Typically the initial distribution is different from the invariant distribution of the Markov chain. As a result the initial part of the chain may consist of states that are not typical for the invariant distribution leading to significantly biased estimates if the Markov chain is not long enough. Typically the solution is to discard the initial part of the chain as a so-called *burn-in*. The hope is that after the burn-in the remaining

part of the chain is sufficiently close to equilibrium that any remaining bias is negligible. The key question is how to choose the length of the burn-in.

The most common way to determine the length of the burn-in is by observing a plot of $k(X_i)$ versus i for one or more real valued functions k. Such plots are referred to as trace plots or times series plots. A natural choice of k would be any sufficient statistics for the target density. The burn-in is then chosen to be a length of time after which the trace plots seems to have reached equilibrium. As the initial state may influence the appropriate length of the burn-in one can start a number of Markov chains in different states and let the burn-in be the time needed before the trace plots associated with the different starting states qualitatively look the same. A recent review of related techniques and topics is given in Robert & Casella (1999).

The length of the burn-in is obviously related to the rate of convergence for the Markov chain in question. Another related matter is mixing properties. In words mixing is a question of how easy the Markov chain is moving around in the state space. This is typically measured by observing estimates of the lag m auto-correlations $\rho^{(m)}$. If $\rho^{(m)}$ tends to zero fast or slow as m increases the Markov chain is said to be fast or slow mixing, respectively.

In many practical situations mixing can be improved by tuning the proposal distributions. When tuning proposal distributions one should note that high acceptance probabilities may indicate that the proposal distributions are too concentrated near the current state implying that the Markov chain is barely moving. Low acceptance probabilities imply that the chain may not move at all for long periods of time. Hence, both high and low acceptance probabilities may be signs of slow mixing. For a Metropolis algorithm theoretical considerations by Roberts, Gelman & Gilks (1997) suggest that an acceptance probability around 0.25 is optimal.

In some situations the target distribution may be of such a nature that any Markov chain is inherently slow mixing. Typically this is because the target distribution is very concentrated in one or more regions of the state space. One technique dealing with this is *simulated tempering* (Marinari & Parisi 1992, Geyer & Thompson 1995). Assume that we have a family of distributions with densities $f_\tau(x|\theta)$, $\tau \in \{0, 1, \ldots, k\}$, chosen so that for increasing τ the densities are less concentrated leading to improved mixing properties. Using terminology from statistical physics we refer to τ as the temperature with the coldest distribution being the target distribution. The basic idea is then to combine transitions in the state space with transitions between distributions at different temperatures. Specifically this is done by specifying a so-called pseudo prior $\pi(\tau)$ for τ and then sample (x, τ) from the joint distribution $f_\tau(x|\theta)\pi(\tau)$ using a Metropolis-Hastings algorithm where the temperature is an additional component. Given the resulting Markov chain $\{(X_i, \tau_i) : i \in \mathbb{N}_0\}$ we use the subsample $\{X_i : \tau_i = 0\}$ as an approximate sample from the target distribution.

A simple way of reducing auto-correlation is *subsampling*. Given a Markov chain $\{X_i : i \in \mathbb{N}_0\}$ this is done by only making use of $\{X_{mi} : i \in \mathbb{N}_0\}$ for $m \in \mathbb{N}$. In addition to reduction of auto-correlation subsampling may be motivated by limited storage. As subsampling is in a sense throwing away information one

should take into consideration the time it takes to generate a sample compared to the time it takes to use it when deciding on the spacing m. For more details on such considerations, see Geyer (1992, 1999).

1.6 Perfect Simulation

The area of perfect simulation has gained much attention since the seminal work by Propp & Wilson (1996). An attention which has stimulated the development and application of new and known methods for perfect simulation. A simulation algorithm is said to be exact if it with probability one in finite time return a realisation from the target distribution. In practice such algorithms will not be exact due to the fact that random number generators always have defects and computers have limited precision. As such the output produced is in general no longer exact. On the other hand the output is perfect in the sense that it is not possible to do better. Hence the name *perfect simulation* (Kendall 1998, Kendall & Møller 2000). In addition to problems with random number generators and computer precision a given algorithm for perfect simulation may be infeasible for practical purposes because of limited resources, e.g. available time or computer storage.

With the advert of the perfect simulation algorithms considered in this section as well as a range of other recent perfect simulation algorithms it has become possible to generate i.i.d. samples from target distributions where one earlier had to rely on approximative method, e.g. based on MCMC. In situations where it is infeasible to produce perfect i.i.d. samples we may return to classical MCMC methods replacing the burn-in with a perfect simulation. This approach renders any discussion of initial distribution and length of burn-in irrelevant. It may be argued that replacing a burn-in with perfect simulation is too time consuming but then so is determining a proper burn-in. In fact one could interpret a perfect sampler as a device for automatically choosing a proper burn-in.

In some situations perfect simulation may become so time or memory consuming that it becomes infeasible to use even a single perfect simulation. In such situations perfect simulation may serve as a tool for evaluating the quality of any approximative method. This could be done by simply comparing results of using perfect simulation with results of using the approximative sampling scheme, e.g. the consequence of replacing an initial perfect simulation with a burn-in. It may be necessary to make this comparison in an easier setup than is required to make perfect simulation feasible. Chapter 4 contains an example of this approach.

The definition of perfect simulation includes simple simulation schemes like inversion and rejection sampling. Usually the term perfect simulation is associated with clever and often quite complicated algorithms for sampling — typically situations where the only feasible alternatives are approximative methods. In this section we briefly review Propp-Wilson's CFTP algorithm with and without assumption about monotonicity as well as Wilson's read-once algorithm. These algorithms are all examples of perfect simulation algorithms based on coupled

Markov chains.

In this section we assume that it is possible to construct a Markov chain $\{X_i : i \in \mathbb{N}_0\}$ with a given invariant distribution using a so-called *stochastic recursive sequence* (SRS). This is done by setting $X_i = \phi(X_{i-1}, R_i)$, $i \in \mathbb{N}$, where the so-called *update function* ϕ is a deterministic function and R_i, $i \in \mathbb{N}$, are i.i.d. random variables of possibly varying dimension. Compared to specifying a Markov chain in terms of a transition kernel a construction using an SRS is closer to the way a Markov chain is implemented in a computer.

A simple example of an SRS construction is a random walk on $\Omega = \{1, 2, \ldots, k\}$ with reflecting barriers at 1 and k. Let the update function be given by $X_i = \phi(X_{i-1}, R_i) = \min\{\max\{X_{i-1} + R_i, 1\}, k\}$ and R_i is uniformly distributed on $\{-1, 1\}$. It is easy to show that the equilibrium distribution of the resulting Markov chain is uniform on Ω.

In the following we consider algorithms for perfect simulation where we allow negative time. In each algorithm below the coupling is done by constructing each chain using the same update function and the same realisation of the random variables $\{R_i : i \in \mathbb{Z}\}$ but starting in different states. The different chains are coupled by using the same update function and random variable $\{R_i : i \in \mathbb{Z}\}$, where $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$. This leads to the following notation. For all $m \in \mathbb{Z}$ and $n \geq m$

$$X_n^m(x) = \phi(\phi(\ldots \phi(x, R_{m+1}), \ldots, R_{n-1})R_n)$$

denotes the state of a Markov chain at time n started at time m in state x generated using the SRS construction.

1.6.1 Propp & Wilson's CFTP

Assume that the target distribution Π is defined on a finite state space Ω and that it is possible to generate a uniformly ergodic Markov chain with Π as its invariant distribution using the SRS construction above. We can then simulate perfect samples from Π using Propp-Wilson's coupling from the past (CFTP) algorithm as follows.

Initially choose starting times $\{T_i \in \mathbb{N} : i \in \mathbb{N}\}$ so that $T_i > T_j$ for $i > j$. Then generate coupled chains started at time $-T_1$ in all states of Ω and find their state at time zero, i.e. given random variables $R_{-T_1+1}, \ldots, R_0$ evaluate $X_0^{-T_1}(x)$ for all $x \in \Omega$. If $X_0^{-T_1}(x) = X_0^{-T_1}(y)$ for all $x, y \in \Omega$ we say that the chains have coalesced at time zero and the common state at time zero is then a sample from Π.

If the chains have *not* coalesced at time zero generate random variables $R_{-T_2+1}, \ldots, R_{-T_1}$, reuse $R_{-T_1+1}, \ldots, R_0$ and evaluate $X_0^{-T_2}(x)$ for all $x \in \Omega$. If $X_0^{-T_2}(x) = X_0^{-T_2}(y)$ for all $x, y \in \Omega$ then the common state at time zero is a sample from Π. This procedure of starting further and further back in time is repeated until all chains have coalesced at time zero, i.e. when $X_0^{-T_i}(x) = X_0^{-T_i}(y)$ for all $x, y \in \Omega$ for some $i \in \mathbb{N}$. The common state is then a sample from Π. Propp & Wilson (1996) note that a doubling scheme for the starting times $T_i = T_0 2^{i-1}$, $T_0, i \in \mathbb{N}$, is near optimal. It is important to note that random

variables R_i are reused and hence either have to be stored or regenerated later. Replacing old random variables with new ones for each starting time result in biased samples.

The CFTP algorithm can be summarised as follows where we initially set $i = 1$ and $T_0 = 0$.

1. Generate $R_{-T_i+1}, \dots R_{-T_{i-1}}$.
2. Evaluate $X_0^{-T_i}(x)$ for all $x \in \Omega$.
3. If $X_0^{-T_i}(x) = X_0^{-T_i}(y)$ for all $x, y \in \Omega$ then return the common state. Otherwise increase i by one and go to step 1.

With probability one this algorithm terminates in finite time and returns a sample from the target distribution.

To make this a little more precise define $\tilde{T} = \inf\{T_i : i \in \mathbb{N}, X_0^{-T_i}(x) = X_0^{-T_i}(y) \text{ for all } x, y \in \Omega\}$ and let $\tilde{x} = X_0^{-\tilde{T}}(x)$ for any $x \in \Omega$. Then uniform ergodicity is required to ensure $\mathbb{P}(\tilde{T} < \infty) = 1$. The need for uniform ergodicity is a clear drawback of the CFTP algorithm. Further, by construction $X_0^{-t}(x) = \tilde{x}$ for all $x \in \Omega$ and $t \geq \tilde{T}$. It is then intuitively clear that a chain started in the infinite past and constructed using the SRS construction is also in state $\tilde{x}$ at time zero and hence $\tilde{x}$ is a sample from Π. For a formal proof and more details, see Propp & Wilson (1996).

At this point it may be natural to ask why not start at time zero and go forward in time until all chains have coalesced and use the common state achieved in this fashion as the sample. This would eliminate the storage problem related to Propp-Wilson's CFTP algorithm and be much simpler to implement. This appealing approach unfortunately results in biased samples. In the random walk example above this procedure would only return 1 and k as samples. Aborting the CFTP algorithm whenever the starting time gets below a given threshold is another way of introducing bias — the so-called user impatience bias. These problems and a whole list of other issues related to CFTP are commented on in Wilson (2000b).

1.6.2 Propp & Wilson's monotone CFTP

Even for moderate sized state spaces it becomes too time consuming and hence infeasible to check that chains started in all possible states of Ω have coalesced at time zero. This problem may be avoided if it is possible to specify a partial ordering $\prec$ on the state space Ω so that the update function is monotone w.r.t. $\prec$ in the sense that $\phi(x, \cdot) \prec \phi(y, \cdot)$ whenever $x \prec y$, $x, y \in \Omega$. Further, assume that Ω contains unique maximal and minimal states, $x_{\min}$ and $x_{\max}$, so that $x_{\min} \prec x \prec x_{\max}$ for all $x \in \Omega$ and define lower and upper processes $L_n^m = X_n^m(x_{\min})$ and $U_n^m = X_n^m(x_{\max})$, respectively. Then any chain is sandwiched in between the upper and lower process in the sense that $L_n^m \prec X_n^m(x) \prec U_n^m$ for any $x \in \Omega$. Propp-Wilson's monotone CFTP algorithm then consist in replacing step 2 in the CFTP algorithm of Section 1.6.1 by evaluating only

$U_0^{-T_i}$ and $L_0^{-T_i}$. In step 3 of the algorithm it suffice to check if $U_0^{-T_i} = L_0^{-T_i}$ as this, by the sandwiching property, implies $X_0^{-T_i}(x) = U_0^{-T_i}$ for all $x \in \Omega$. As the basic algorithm is the same as in Section 1.6.1 the output is distributed according to Π and defining $-\tilde{T} = \sup\{-T_i : i \in \mathbb{N}, U_0^{-T_i} = L_0^{-T_i}\}$ we have $\mathbb{P}(\tilde{T} < \infty) = 1$.

If the update function is anti-monotone, i.e. $\phi(y, \cdot) \prec \phi(x, \cdot)$ whenever $x \prec y$ then we can still construct upper and lower processes. This is done using the crossover trick given by Kendall (1998) where $L_n^m = \phi(U_{n-1}^m, R_n)$ and $U_n^m = \phi(L_{n-1}^m, R_n)$, $n > m$. Then $L_n^m \prec X_n^m(x) \prec U_n^m$ still holds for all $x \in \Omega$. Even in some cases of infinite state space is it possible to apply monotone CFTP, see e.g. Häggström, van Lieshout & Møller (1999).

The update function used for the random walk example with reflecting barriers is monotone with respect to $\leq$ as the partial ordering. An interactive illustration of this example of Propp-Wilson's monotone CFTP algorithm can be found at `http://probability.ca/jeff/java/cftp.html`.

1.6.3 Wilson's read-once algorithm

The term read-once refers to algorithms where each R_i is used only once unlike e.g. the Propp-Wilson CFTP algorithm. In this section we consider the read-once algorithm for perfect simulation given by Wilson (2000a).

The key component in Wilson's read-once algorithm is i.i.d. random maps

$$F_i(\cdot) = \phi(\phi(\dots\phi(\cdot, R_{i,1}), \dots, R_{i,M-1}), R_{i,M}), \quad i \in \mathbb{N}$$

where ϕ is an update function and $R_{i,1}, \dots, R_{i,M}$ are i.i.d. random variables independent of i defined as in the SRS construction. We say that F_i is coalescent if $F_i(x) = F_i(y)$ for all $x, y \in \Omega$. To obtain a true read-once algorithm we further need that coalescence can be checked on the fly when generating a random map. Assume that we can find M such that $p = \mathbb{P}(F_i \text{ coalescent}) > 0$. Wilson (2000a) suggest that M is chosen so that $p > 1/2$ which may take some experimentation.

Consider the following construction. Generate i.i.d. random maps $F_1, F_2, \dots$ until F_T is coalescent and set $\tilde{x} = F_1(F_2(\dots F_T(x)\dots))$ for any $x \in \Omega$. Then $\tilde{x}$ is a sample from Π. To see this is true one may think of this construction as the Propp-Wilson CFTP algorithm with starting times $T_i = iM$, $i \in \mathbb{N}$, and coalescence being checked by checking for coalescence separately in each "block" of M updates. As the random maps are applied in reversed order compared to how they are generated this construction is not a read-once algorithm. Noting that T is geometrically distributed with mean $(1-p)/p$ a read-once algorithm is obtained by reversing the generation of random maps as follows.

Assume we can draw T from a geometric distribution with mean $(1-p)/p$. Then generate one random map F_1 condition to be coalescent and generate i.i.d. random maps $F_2, \dots, F_T$ all conditioned not to be coalescent. Then $\tilde{x} = F_T(F_{T-1}(\dots F_1(x)\dots))$ is a sample from the target distribution for any $x \in \Omega$. As the random maps are now applied as they are generated the method is a read-once perfect sampler. Only problem is that p is in general unknown. This problem is solved in the Wilson's read-once algorithm.

1. Generate i.i.d. random maps until a coalescent one is generated and name it F_1.
2. Then generate i.i.d. random maps $F_2, F_3, \ldots$ until F_T is coalescent.
3. Return $\tilde{x} = F_{T-1}(F_{T-2}(\ldots F_1(x)\ldots))$ for any $x \in \Omega$

By construction $T-1$ is geometrically distributed with mean $(1-p)/p$ as required and hence $\tilde{x}$ is a sample from Π.

As the random maps are i.i.d. the unused random map F_T may be used as the initial coalescent random map when generating the next perfect sample. Hence, to generate N realisations from Π we need to generate i.i.d. random maps until we have generated $N+1$ that are coalescent.

1.6.4 Other perfect simulation algorithms

There exists a whole range of different perfect simulation schemes. Fill (1998) introduces an algorithm which combines coupling and rejection sampling resulting in an algorithm which does not suffer from user impatience bias. Møller & Nicholls (1999) combine simulated tempering with CFTP for perfect simulation, an idea which is further developed by Brooks, Fan & Rosenthal (2002). A comprehensive list of references to literature on perfect simulation can be found at David B. Wilson's homepage `http://www.dbwilson.com/exact/`.

1.7 Ratios of unknown normalising constants

Ratios of unknown normalising constants appear in many statistical problems. In this thesis a recurring situation is the Bayesian setup: Given likelihood $\pi(x|\theta) = c(\theta)^{-1} f(x|\theta)$ and prior $\pi(\theta)$ we then want to sample from the posterior $\pi(\theta|x) = \pi(\theta)\pi(x|\theta)$ using a Metropolis-Hastings algorithm. Unlike when we sample the likelihood the Hastings ratio now includes a ratio of normalising constants that do not in general cancel. As the normalising constant is typically unknown we need a method for evaluating (or at least estimating) a ratio of unknown normalising constants.

In this section we consider methods for estimating ratios of unknown normalising constants based on importance, bridge and path sampling. These methods also serves as examples of what easily becomes computer intensive methods. Even if evaluating a ratio of unnormalised likelihoods is easy a simulation based estimation of the ratio of the unknown normalising constants may slow everything down considerably. The last simulation of a posterior sample in Chapter 4 took around a week to complete where most time was spend on estimating ratios of unknown normalising constants.

In Chapter 5 we introduce an auxiliary variable methods which avoids estimating the ratio of unknown normalising constants in the Hastings ratio. This in a sense makes the Metropolis-Hastings algorithm exact. The eliminated uncertainty comes at a price which is illustrated in a simulation study in Chapter 5 and in an application in Chapter 6.

1.7.1 Importance and bridge sampling

Consider a density $f(x|\theta)$ with unknown normalising constant $c(\theta)$. Assume that parameters θ_0 and θ_1 have the property that $f(x|\theta_1) > 0$ implies $f(x|\theta_0) > 0$ for all $x \in \Omega$. The ratio of unknown normalising constants is then given by a special case of the *importance sampling formula*

$$\frac{c(\theta_1)}{c(\theta_0)} = \mathbb{E}_{\theta_0}\left[\frac{f(X|\theta_1)}{f(X|\theta_0)}\right] \tag{1.5}$$

where $\mathbb{E}_{\theta_0}$ denotes expectation with respect to $f(\cdot|\theta_0)$. The ratio of unknown normalising constants can then be approximated by

$$\frac{1}{N}\sum_{i=1}^{N}\frac{f(X_i|\theta_1)}{f(X_i|\theta_0)} \tag{1.6}$$

where $X_1, \ldots, X_N$ are either an i.i.d. sample from $f(\cdot|\theta_0)$ or a Markov chain with invariant distribution $f(\cdot|\theta_0)$. In practice θ_0 and θ_1 have to be close for the approximation (1.6) to work well, see Geyer (1999) for more details.

Assume that θ_0 and θ_1 are so far apart that (1.6) is not a feasible approximation. A natural extension of (1.5) is then to introduce an intermediate distribution as a bridge between the two distributions. Assume that $\theta_{.5}$ is a parameter "between" θ_0 and θ_1 with the property that $f(x|\theta_{.5}) > 0$ implies $f(x|\theta_0) > 0$ and $f(x|\theta_1) > 0$ for all $x \in \Omega$. Then

$$\frac{c(\theta_1)}{c(\theta_0)} = \frac{c(\theta_1)}{c(\theta_{.5})}\frac{c(\theta_{.5})}{c(\theta_0)} = \frac{\mathbb{E}_{\theta_0}[f(X|\theta_{.5})/f(X|\theta_0)]}{\mathbb{E}_{\theta_1}[f(X|\theta_{.5})/f(X|\theta_1)]}, \tag{1.7}$$

is a special case of so-called *bridge sampling* introduced by Meng & Wong (1996). Like (1.5) it can be approximated by replacing the two expectations by Monte Carlo using a sample from $f(x|\theta_0)$ and $f(x|\theta_1)$, respectively. This approach can be extended in an obvious way to any number of intermediate distributions.

1.7.2 Path sampling

Path sampling is related to what in statistical physics is known as thermodynamic integration. In statistics it has been applied in statistics in several special guises. Interestingly an early application of path sampling by Ogata & Tanemura (1984) is in fact to point processes. Gelman & Meng (1998) give a general description of path sampling in statistics and its relation to importance and bridge sampling.

Assume that an unnormalised density $\pi(x|\theta) = c_\theta^{-1} f(x|\theta)$ is specified by a single parameter $\theta \in \Theta$. Further, assume that $\theta : [0,1] \mapsto \Theta$ is a differentiable path in the parameter space Θ, where $\theta(0) = \theta_0$ and $\theta(1) = \theta_1$. The *path sampling identity* is then given by

$$\ln(c(\theta_1)/c(\theta_0)) = \int_0^1 \mathbb{E}_{\theta(s)}[U(x,\theta(s))]\theta'(s)\mathrm{d}s, \tag{1.8}$$

where $\theta'(s) = \mathrm{d}\theta(s)/\mathrm{d}s$ and $U(x,\theta) = \mathrm{d}\ln f(x|\theta)/\mathrm{d}\theta$. The derivation of the path sampling identity (1.8) is rather straight forward under the assumption that it is legitimate to interchange differentiation and integration of $f(x|\theta)$. Gelman & Meng (1998) give a theoretical treatment of how to choose the optimal path. Furthermore, compared to importance and bridge sampling estimates based on path sampling are expected to be more stable as they are calculated on the log scale.

Gelman & Meng (1998) consider two implementations of the path sampling identity. The first method is based on estimating the expectation in (1.8) at a number of sites along the path with expectations typically estimated by MCMC. The estimated expectations are then combined with a Riemann approximation of the path integral (1.8). We have applied this approach in Chapter 4 in a point process setting.

The second implementation considered by Gelman & Meng (1998) is related to simulated tempering. As in simulated tempering we introduce a so-called pseudo prior $\pi(s)$. The idea is then to sample (x,s) from the joint unnormalised density $f(x|\theta(s))\pi(s)$ using e.g. Metropolis-Hastings. Assume that $(x_0,s_0),\ldots,(x_N,s_N)$ is such a sample where the sample has been sorted so that $s_i \leq s_{i+1}$ for $i = 0,\ldots,N-1$ then

$$\ln\left[\frac{c(\theta_1)}{c(\theta_0)}\right] \approx \sum_{i=1}^{N} \frac{1}{2}(s_i - s_{i-1})\big((U(x_{i-1},\theta(s_{i-1}))\theta'(s_{i-1}) + U(x_i,\theta(s_i))\theta'(s_i)\big).$$

Compared to the grid approach the advantage of this method is that only one Markov chain is needed, i.e. only one burn-in or initial perfect simulation is needed.

Gelman & Meng (1998) conclude that in practice it is difficult to calculate the optimal variance reducing pseudo prior and instead aim for the pseudo prior leading to a uniform marginal distribution of s. This requirement is equivalent to $\pi(s) \propto c_{\theta(s)}^{-1}$, i.e. the unknown normalising constant inverted. Gelman & Meng (1998) propose an iterative procedure for approximating the pseudo prior. In connection with Chapter 6 numerical experiment have been carried out to assess the difference in performance for the two implementations of path sampling. This was done in the setting of the Metropolis-Hastings algorithm for posterior sampling considered in Chapter 6. These experiments indicated that the simulated tempering like approach is more time consuming than the grid type approach for comparable precision. Further, the experiments showed that importance sampling would produce a sufficiently good approximation of the pseudo prior making any iterative updates unnecessary. This may be explained by the fact that the proposal distributions in Chapter 6 are concentrated near the current state.

1.8 Point processes

Informally a point process is a random countable subset of some region S. For simplicity we assume for the remainder of this chapter that $S \subseteq \mathbb{R}^d$ which is sufficient for most practical purposes. An introduction to general point process theory is given in (Daley & Vere-Jones 1988). Møller & Waagepetersen (2003) give a less technical introduction to point processes aimed at a broader audience.

By nature point processes are stochastic models for point patterns. Point patterns arise in many situations typically as the locations of objects. Examples include the positions of trees in a forest or the locations of cells in tissue. Examples of the latter are considered in Chapters 4 and 6. In some situations more information about individual objects is available in addition to location. Such additional information may be included in the shape of marked point processes considered briefly in Section 1.8.1.

The simplest point process is the Poisson process considered in Section 1.8.2. The Poisson process can be interpreted as a model for no interaction between points and as such it is usually not very interesting as a model. On the other hand the Poisson process is mathematically tractable making it interesting as a reference model for more complicated models. Cox processes considered in Section 1.8.3 mainly serve as models for clustered point patterns. Regular point patterns indicate repulsion between points which is can be modelled by pairwise interaction point processes considered in Section 1.8.4. Pairwise interaction point processes are a sub-class of so-called Markov point processes. For a comprehensive introduction to Markov point processes, see van Lieshout (2000).

Before we continue the treatment of point processes we need some notation. Let $\mathcal{B}$ denote the Borel sets of S and $\mathcal{B}_0$ the bounded Borel sets of S. The state space is the set of locally finite point configurations $\Omega = \{x \subset S : n(x \cap B) < \infty, \forall B \in \mathcal{B}_0\}$ where x is a point configuration and $n(x)$ denotes the cardinality of x. For Ω define the σ-algebra $\mathcal{F} = \sigma(\{\{x \in \Omega : n(x \cap B) = m\} : m \in \mathbb{N}_0, B \in \mathcal{B}_0\})$. The formal definition of a point process is then a measurable function X from some probability space to the measure space $(\Omega, \mathcal{F})$.

1.8.1 Marked point processes

Assume that X is a point process where each point $\xi \in X$ is associated with a random mark $m_\xi \in \mathcal{M}$ where $\mathcal{M}$ is the so-called mark space. Then $Y = \{(\xi, m_\xi) : \xi \in X\}$ is a marked point process.

A simple example is a so-called *disc process* where the point process X specifies the centres of discs and the associated marks in $\mathcal{M} = [0, \infty)$ correspond to the radii of the associated discs. In applications this may be a suitable model for a forest stand where the radii could refer to the stem radii. Another example is so-called *multi-type* point processes where the mark space is finite and different marks indicate different types of points. In applications one may think of different types of cells in tissue. If each mark m_ξ is a random closed set, e.g. a line segment, an ellipsis or a polygon located at ξ the result is a *germ-grain* model where the points of X constitutes the germs and the grains are given by

the marks.

In Chapter 4 we interpret the configuration of a random step function as a marked point process on a line segment where the points specify the locations of the steps and each marks correspond to the level of that step.

1.8.2 Binomial and Poisson point processes

Assume that f is a density function on $B \in \mathcal{B}$, $|B| > 0$. A *binomial point process* on B with density f conditioned to have m points consists of m i.i.d. points with density f. A simple and often recurring situation is when $f = 1/|B|$ is the uniform distribution.

Let $\rho : S \mapsto [0, \infty)$ be the so-called *intensity function* which is locally integrable and define the *intensity measure* $\mu(B) = \int_B \rho(\xi) \mathrm{d}\xi$ for all $B \in \mathcal{B}$. A Poisson point process X on S with intensity measure μ is then a point process satisfying **1)** $n(X \cap B)$ is Poisson distributed with mean $\mu(B)$ for all $B \in \mathcal{B}$ with $\mu(B) < \infty$ and **2)** for all $m \in \mathbb{N}$ and $B \in \mathcal{B}$ with $0 < \mu(B) < \infty$, conditional on $n(X \cap B) = m$ then $X \cap B$ is a binomial process with density $\rho(\xi)/\mu(B)$ conditioned to have m points.

If $\rho(\xi) = \rho_0$ is constant the resulting Poisson process is homogeneous with rate ρ_0. Otherwise the Poisson process is said to be inhomogeneous.

For a proof of existence and uniqueness of the Poisson point process together with proofs of the following properties for the Poisson process see Møller & Waagepetersen (2003). If X is Poisson point process then $X \cap B_1, X \cap B_2, \ldots$ are independent for disjoint subsets $B_1, B_2, \ldots \in \mathcal{B}$. This property is known as the independent scatter property or complete spatial randomness. Let $X_1, X_2, \ldots$ be Poisson processes where X_i has intensity function ρ_i. Assuming that $\rho = \sum_i \rho_i$ is locally integrable then $X = \cup_i X_i$ is with probability one a disjoint union and X is a Poisson process with intensity function ρ. Poisson processes are in general mathematically tractable making it possible to derive a wide range of further theoretical properties, see e.g. Daley & Vere-Jones (1988), Kingman (1993) and Stoyan, Kendall & Mecke (1995).

1.8.3 Cox processes

A Poisson process is typically an unsatisfactory model for most point patterns of interest. Cox processes constitutes an interesting class of point processes especially suited for clustered data. Assume that $\rho = \{\rho(\xi) > 0 : \xi \in S\}$ is a non-negative random field on S which is locally integrable with probability one. If X conditional on ρ is a Poisson process with intensity function ρ then X is a Cox process driven by ρ. Such processes were originally introduced in a time series setting by Cox (1955) who referred to them as doubly stochastic.

For a simple example let λ be a positive random variable and conditional on λ let X be a Poisson process with constant intensity function $\rho(\xi) = \lambda$. Then X is a so-called mixed Poisson process. For more details on Cox processes, see Daley & Vere-Jones (1988), Stoyan et al. (1995), Møller & Waagepetersen (2002) and Møller & Waagepetersen (2003).

So-called shot noise Cox processes (SNCP) are a large class of Cox processes driven by random intensity functions of the form $\rho(\xi) = \sum_i \lambda_i \mathcal{K}(\xi; c_i)$. Here $(c_1, \lambda_1), (c_2, \lambda_2), \ldots$ are the points of a Poisson process on $S \times [0, \infty)$ specified by a locally integrable intensity function and $\mathcal{K}(\cdot;\cdot)$ is a kernel, i.e. $\mathcal{K}(\cdot; c)$ is a probability density for all $c \in S$. Conditional on ρ a SNCP may be interpreted as the union of inhomogeneous Poisson processes each with intensity function $\lambda_i \mathcal{K}(\cdot; c_i)$. For more details on SNCPs, see e.g. Møller (2003).

In Chapter 6 we analyse an inhomogeneous point pattern using what could be called a generalised SNCP as the prior distribution. Generalised in the sense that the random intensity measure is unchanged but conditional on the random intensity measure X is not a Poisson process but a pairwise interaction process, a class of point processes considered below.

1.8.4 Point processes with a density

Assume that X is a point process in S with unnormalised density f with respect to a unit rate Poisson process on S. Then

$$\mathbb{P}(X \in F) = c^{-1} \sum_{n=0}^{\infty} \frac{e^{-|S|}}{n!} \int_S \cdots \int_S \mathbf{1}[\{x_1, x_2, \ldots, x_m\} \in F] \times f(\{x_1, x_2, \ldots, x_n\}) \mathrm{d}x_1 \cdots \mathrm{d}x_n,$$

where $F \in \mathcal{F}$ and c is the normalising constant which typically is unknown as it involves evaluating a mathematically intractable high dimensional integral.

The *Papangelou conditional intensity* for a point process is defined by $\lambda(x, \xi) = f(x \cup \xi)/f(x)$, $x \in \Omega$, $\xi \in S \backslash x$ and taking $a/0 = 0$ for $a \geq 0$ (Kallenberg 1984). A point process is said to be *attractive* if $\lambda(x, \xi) \leq \lambda(y, \xi)$ whenever $x \subseteq y$ and *repulsive* if $\lambda(x, \xi) \geq \lambda(y, \xi)$ whenever $x \subseteq y$.

A density f is *locally stable* if there exists a function K with $\int_S K(\xi) \mathrm{d}\xi < \infty$ so that $f(x \cup \xi) \leq f(x) K(\xi)$ for all $x \in \Omega$ and $\xi \in S \backslash x$. Local stability implies *Ruelle stability*. A density is Ruelle stable if there exist constant α and function K with $\int_S K(\xi) < \infty$ so that $f(x) \leq \alpha \prod_{\xi \in x} K(\xi)$ for all $x \in \Omega$. A Ruelle stable density is integrable w.r.t. a unit rate Poisson process, i.e. $c < \infty$.

In many practical application a useful class of point processes is the *pairwise interaction point processes* that have densities of the form

$$f(x) \propto \prod_{\xi \in x} \beta(\xi) \prod_{\{\xi, \eta\} \subseteq x : \xi \neq \eta} \varphi(\{\xi, \eta\}), \tag{1.9}$$

where the first order term $\beta : S \mapsto [0, \infty)$ is known as the chemical activity in statistical physics. The second order term $\varphi : S \times S \mapsto [0, \infty)$ is referred to as the *interaction function*. If $\varphi \leq 1$ the process is repulsive and if in addition $\int_S \beta(\xi) \mathrm{d}\xi < \infty$ the process is locally stable and hence well defined. If $\varphi \geq 1$ the process is attractive and not in general well-defined unless we condition on the number of points.

Define the *interaction range* as $R = \sup_{\{\xi,\eta\}\subseteq S}\{\|\xi - \eta\| : \varphi(\{\xi,\eta\}) \neq 1\}$ where $\|\cdot\|$ denotes Euclidean distance. Then with f given by (1.9) the Papangelou conditional intensity only depends on $X \cap \text{Ball}(\xi, R)$ where R is the interaction range and $\text{Ball}(\xi, R)$ denotes a closed ball centres at ξ with radius R.

A simple example of a widely used pairwise interaction point process is the Strauss process (Strauss 1975) where $\beta(\xi) \equiv \beta$ is constant and $\varphi(\{\xi,\eta\}) = \gamma^{\mathbf{1}[\|\xi-\eta\|\leq R]}$ with $0^0 = 1$. Here $0 \leq \gamma \leq 1$ is an interaction parameter and $R \geq 0$ is equivalent to the interaction range. Let $s_R(x) = \sum_{\{\xi,\eta\}\subseteq x:\xi\neq\eta} \mathbf{1}[\|\xi - \eta\| \leq R]$ denote the number of pairs of point in x with inter point distance less than R. The Strauss process then has unnormalised density $\beta^{n(x)}\gamma^{s_R(x)}$ w.r.t. a unit rate Poisson process. The Strauss process is a recurring example through out this thesis. In Chapters 2 and 3 we demonstrate and compare different perfect simulations methods by applying them to a Strauss process. In Chapter 4 we approximate the unknown normalising constant for a Strauss process on a grid of parameter values. In Chapter 5 we study the feasibility of an auxiliary variable method through a Bayesian analysis of the parameters β and γ of a Strauss process.

Other pairwise interaction point processes are considered in this thesis. In Chapter 4 we perform a non-parametric Bayesian analysis of the interaction function using a multi-scale process which is a simple extension of the Strauss process introduced by Penttinen (1984). In Chapter 4 we also consider an interaction function applied by Diggle & Gratton (1984) for the analysis of the data considered in the same chapter. In Chapter 6 we consider a pairwise interaction point process with an inhomogeneous first order term.

1.9 Point process simulation

It is usually only quite simple point processes that can be simulated perfectly without relying on either rejection sampling, coupling or other novel techniques. One obvious exception is point processes that are specified by how they are simulated, e.g. simple sequential inhibition (Diggle, Besag & Gleaves 1976). In this section we consider simulation of Poisson and Cox processes and demonstrate that perfect simulation of point processes using rejection sampling is infeasible even for point processes with a moderate degree of interaction.

We briefly review how point processes are simulated using a Metropolis-Hastings algorithm in both the fixed and variable dimensional cases equivalent to simulating a point process where we have conditioned respectively not conditioned on the number of points.

An alternative to Metropolis-Hastings simulation of point processes is continuous time birth-and-death processes. Chapter 2 contains a description of how locally stable point processes can be generated using a continuous time birth-and-death process by a coupling to a simple continuous time jump process on the non-negative integers. This coupling construction further serves as the back bone of a dominated coupling from the past (dominated CFTP) algorithm for

perfect simulation of locally stable point processes (Kendall 1998, Kendall & Møller 2000) reviewed in Chapters 2 and 3, and applied through out this thesis. Briefly dominated coupling from the past works by starting a reference chain at time zero and then run it backwards in time. This is followed by a coupling construction forward in time similar to ordinary CFTP except it includes a coupling to the reference process. Using dominated CFTP as considered by Kendall & Møller (2000) for simulating locally stable point processes the reference processes stochastically dominates the coupled processes run forward in time, hence the name. In addition to dominated CFTP based on spatial birth-and-death processes Kendall & Møller (2000) present a version based on Metropolis-Hastings updates.

Some times dominated CFTP is referred to as *coupling into and from the past* (CIAFTP) as the reference chain does not have to be dominating. For an example of this, see e.g. Kendall & Thönnes (1999).

In Chapter 3 we review a perfect simulation algorithm using clans of ancestors based on spatial birth-and-death processes. This algorithm is further compared to dominated CFTP in a simulation study.

The last simulation algorithm considered in this section is a version of Wilson's read-once algorithm for perfect simulation of locally stable point processes.

1.9.1 Poisson process simulation

Simulation of a Poisson process follows immediately from the definition. Assume that X is a Poisson process on S with intensity function ρ and corresponding intensity measure μ. Then $X \cap B$, $B \in \mathcal{B}$ with $\mu(B) < \infty$ is simulated by first generating $m \in \mathbb{N}_0$ from a Poisson distribution with mean $\mu(B)$. Next generate a binomial process X on B with density $\rho(\xi)/\mu(B)$ conditioned to have m points. This is particularly easy if B is rectangular and ρ is constant. If ρ is non-constant and has upper bound ρ^* an alternative is to generate a homogeneous Poisson process on B with intensity ρ^* followed by an independent thinning where each point $\xi \in X$ is retain with probability $\rho(\xi)/\rho^*$ and otherwise deleted. See e.g. Stoyan et al. (1995) for more details on independent and dependent thinning.

1.9.2 Cox process simulation

Given a realisation of the random intensity function simulating a Cox process is a question of simulating an inhomogeneous Poisson process which in principle is easy. The difficulty is thus typically in the simulation of the random intensity function.

Assuming $S = \mathbb{R}^d$ simulating a SNCP on a bounded subset $W \subset S$ would at first sight be impossible as the random intensity function may require the simulation of a possibly infinite Poisson process on $S \times [0, \infty)$. If the kernel $\mathcal{K}(\cdot, c)$ has bounded support for all $c \in S$ it is possible to simulate the SNCP by simulating the underlying Poisson process on a region $W_{\text{ext}} \times [0, \infty)$ where $W_{\text{ext}} \supseteq W$ is a bounded set. For kernels with unbounded support this procedure of simulating the Poisson process on a bounded set W_{ext} would introduce an

edge effect due to the missing contributions to the random intensity function from kernels $\mathcal{K}(\cdot, c)$ with $c \in S \backslash W_{\text{ext}}$. Fortunately, Brix & Kendall (2002) give a method for perfect simulation of SNCP avoiding any edge effect even for kernels with unbounded support.

1.9.3 Rejection sampling for point processes

In Section 1.2 we considered rejection sampling as a relatively simple way of producing perfect samples provided that a covering density exists from which we can sample. Restricting attention to repulsive pairwise interaction point processes it is easily shown that a pairwise interaction point process with density (1.9) w.r.t. a unit rate Poisson process is covered by the unnormalised density $\prod_{\xi \in x} \beta(\xi)$ of an inhomogeneous Poisson point process. In this setup the acceptance probability is $\alpha(x) = \prod_{\{\xi,\eta\} \subseteq x : \xi \neq \eta} \varphi(\{\xi, \eta\})$. For specificity we consider a simulation study below comparing rejection sampling and dominated CFTP based on a spatial birth-and-death process for a Strauss process. In this specific situation $\alpha(x) = \gamma^{s_R(x)}$, i.e. in the Poisson case ($R = 0$ or $\gamma = 1$) all proposals are accepted.

Some results of the simulation study are presented in the left and centre plot of Figure 1.1 showing the differences in computing times for rejection sampling and dominated CFTP. The figure shows that rejection sampling is inferior to dominated CFTP already at moderate levels of interaction, i.e. when $R > 0$ or $\gamma < 1$. So even if simulating a Poisson process is cheap in terms of computing time and storage this does not compensate for the fact that the Poisson processes is too far from a Strauss process to be useful as a covering density. This fact is further emphasised by the right plot of Figure 1.1 showing the differences in the empirical means of s_R for the Poisson and Strauss processes.

In Chapter 5 we consider a partially ordered Markov model (POMM) point process on on a bounded $S \subset \mathbb{R}^2$ as an approximation of a Strauss point process. Assuming that β is constant we show below that the density of this POMM point process covers the Strauss density and report on some simulation experiments using the POMM point process for rejection sampling. The POMM point process in question has density

$$f_P(x|\beta_P, \gamma_P, R_P) = \exp[-\beta_P \sum_{i=1}^{M} |C_i| \gamma_P^{s_i(x)}] \beta_P^{n(x)} \prod_{i=1}^{M} \gamma_P^{s_i(x) n_j(x)}$$

w.r.t. a unit rate Poisson process on S. Here $\{C_i \subseteq S : i = 1, \ldots, M\}$ constitutes a division of S into disjoint subsets referred to as cells. Furthermore, $n_i(x) = n(x \cap C_i)$ and $s_i(x) = \sum_{j<i} n_j(x) \mathbf{1}[\|\xi_i - \xi_j\| \leq R_P]$ where $\{\xi_i \in C_i : i \in \{1, \ldots, M\}\}$ are prespecified fixed points. For more details see Chapter 5.

To use the POMM point process for rejection sampling we need positive constants β_P, γ_P, R_P and k so that $k f_P(x|\beta_P, \gamma_P, R_P) \geq \beta^{n(x)} \gamma^{s_R(x)}$ for all $x \in \Omega$. Choosing $\beta_P = \beta$ and $\gamma_P = \gamma$ this is achieved if $k = \exp(\beta|S|)$ and R_P is chosen so that $\sum_{i=1}^{M} s_i(x) n_i(x) \leq s(x)$ for all x. The latter is fulfilled if all cells are squares with identical side lengths Δ and $R_P = \max\{R - \sqrt{2}\Delta, 0\}$.

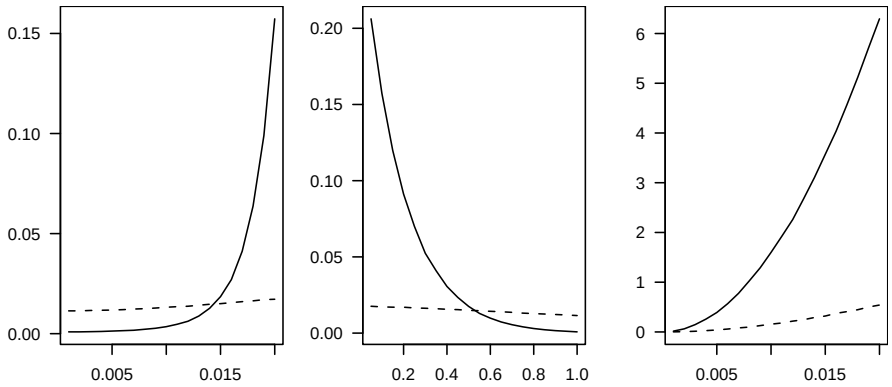

Figure 1.1: Comparison of rejection sampling and dominated CFTP based on spatial birth-and-death process for a Strauss process on the unit square with $\beta = 100$ and varying values of γ and R. Left plot: mean CPU time per sample (in seconds) versus R for the Strauss process with $\gamma = 0.1$ (rejection sampling: solid line; dominated CFTP: dashed line). Centre plot: mean CPU time per sample versus γ for the Strauss process with $R = 0.02$ (rejection sampling: solid line; dominated CFTP: dashed line). Right plot: $\mathbb{E}s_R$ versus R for the Strauss process with $\gamma = 0.1$ (dashed line) and a homogeneous Poisson process on the unit square with intensity $\beta = 100$ (solid line).

Assume that $S = [0, 1]^2$ is the unit square divided into $M = 100 \times 100$ square cells and let $\beta = 100$, $\gamma = 0.1$ and $R = 0.02$. For this particular case using rejection sampling with proposals from respectively a Poisson and a POMM process result in mean acceptance probabilities of 0.0055 and 0.0062. The slight improvement in acceptance probabilities going from Poisson to POMM proposals is ruined by a significant increase in computing times. More precisely, the mean computing times for rejection sampling using Poisson and POMM proposals are $0.15s$ and $5.67s$, respectively. Increasing the number of cells to $M = 200 \times 200$ does not change the picture significantly. In Chapter 5 we conclude that β_P should be less than β for the POMM point process to be a good approximation of the Strauss process. As we have used $\beta_P = \beta$ above the covering density is not the optimal POMM approximation which may be part of the explaination why rejection sampling using POMM point process proposals performs so poorly.

1.9.4 Metropolis-Hastings - conditional case

Assume that we have a point process X on $S \in \mathcal{B}_0$ with density f with respect to a unit rate Poisson process conditioned to have n point. The Metropolis-Hastings sampler for the fixed dimensional case can then be used for sampling. Using the notation of Section 1.4.1 we assume for simplicity that each component

consist of only one point, i.e. $\Omega_i = S$, $i = 1, \ldots, n$ and $\Omega = S^n$. This is equivalent to the situation where only one point is moved at a time. The simplest situation is the Metropolis case where components, i.e. points, are selected uniformly at random and proposals for new points are generated uniformly on S. Generating a Markov chain $\{X_i : i \in \mathbb{N}_0\}$ assume for $i = 0, 1, 2, \ldots$ that $X_i = x$ is the current state. Then

1. Select η uniformly randomly from x.
2. Generate point ξ uniformly on S
3. With probability $\min\{1, f((x\backslash\{\eta\}) \cup \{\xi\})/f(x)\}$ set $X_{i+1} = (x\backslash\{\eta\}) \cup \{\xi\}$ otherwise set $X_{i+1} = x$.

An alternative to the Metropolis algorithm above is Gibbs sampling introduced in the point process setting by Ripley (1977, 1979). In this setting the uniformly distributed point in step 2 is replaced by a point with a density proportional to $f((x\backslash\{\eta\}) \cup \cdot)$. In most cases the nature of f is such that the new point must be simulated by rejection sampling. The latter is usually difficult rendering the appealing "always accept" aspect of the Gibbs sampler some what illusive.

1.9.5 Metropolis-Hastings - unconditional case

Geyer & Møller (1994) give an algorithm for simulating from the unconditional point process. Their algorithm is essentially a special case of the reversible jump algorithm by Green (1995) considered in Section 1.4.2 where the birth or death of a single point at a time are the dimension changing moves. We only consider a basic version on the Geyer-Møller algorithm. For details on a more general setup, see Geyer & Møller (1994).

Generating a Markov chain $\{X_i : i \in \mathbb{N}_0\}$, for $i = 0, 1, 2, \ldots$ assume that the current state is $X_i = x$. Then with probability $1/2$ choose birth and otherwise choose death. In case of birth generate a point ξ uniformly on S and with probability $\min\{1, [f(x \cup \{\xi\})|S|]/[f(x)(n(x) + 1)]\}$ set $X_{i+1} = x \cup \{\xi\}$ and otherwise set $X_{i+1} = x$. In case of death select uniformly at random a point $\eta \in x$ and with probability $\min\{1, [f(x\backslash\{\eta\})n(x)]/[f(x)|S|]\}$ set $X_{i+1} = x\backslash\{\eta\}$ and otherwise set $X_{i+1} = x$.

1.9.6 Dominated CFTP and read-once

Going from CFTP to read-once in Section 1.6.3 basically consisted in introducing appropriate random maps and then rewriting the CFTP algorithm in a clever way. One reason this worked was that the individual update functions were independent. With dominated CFTP it is not possible to follow the same procedure to obtain a read-once algorithm. The reason is that the individual update functions in this case are dependent because of the coupling to the dominating reference process.

Wilson (2000a) gives a general description for how to make a read-once algorithm based on dominated CFTP. However the resulting algorithm suffers from infinite expected running time. As a special case he presents a purpose build read-once algorithm for locally stable point processes which we consider here. This read-once algorithm does not suffer from infinite expected running time.

The read-once algorithm considered below works like the read-once algorithm in Section 1.6.3 by specifying i.i.d. random maps that maps Ω into Ω and are coalescent with positive probability. The random maps do not dependent on their input state and as such we can only assume that the input state is in Ω and hence consists of any number of points that could be anywhere. In the following we refer to these points as unknown. Further, we assume that the target process is a locally stable point process as defined in Section 1.8.4 with $K(\xi) = K$ constant.

Each random map consists of three parts. First a Metropolis-Hastings step that leaves us with an upper bound on the number of unknown points. Secondly, time is set equal zero and a birth-and-death type process is run until all unknown points have died. Finally, upper and lower processes are generated as in dominated CFTP until a prespecified time T. Let X_0 denote the a priori unknown input state of the random map. Further, at any time X denotes the state of the target process, $X^{(u)} \subseteq X$ denotes the set of unknown points in X, D denotes the set of known points, and k denotes the maximal number of unknown points in X. Initially we have $X = X_0$, $D = \emptyset$ and $k = \infty$.

First part of the random map is a Metropolis-Hastings update where the proposal x' is a Poisson process with intensity $2K$. The Hastings ratio is then

$$H(X_0, x') = \frac{f(x')(2K)^{n(X_0)}}{f(X_0)(2K)^{n(x')}} \tag{1.10}$$

and with probability $\alpha(X_0, x') = \min\{1, H(X_0, x')\}$ we set $X = x'$ and otherwise retain the input, i.e. $X = X_0$. Local stability implies a lower bound on the Hastings ratio

$$H(X_0, x') \geq \frac{f(x')(2K)^{n(X_0)}}{f(\emptyset)K^{n(X_0)}(2K)^{n(x')}}. \tag{1.11}$$

Bounding the RHS of (1.11) from below by one we get that the proposal x' is accepted at least whenever $n(X_0) \geq B(x') \equiv \log_2 f(\emptyset) + n(x') + n(x') \log_2 K - \log_2 f(x')$. Using local stability again gives $\log_2 f(x') \leq n(x') \log_2 K + \log_2 f(\emptyset)$ which implies $B(x') \geq n(x')$. Hence, after the initial Metropolis-Hastings update X contains at most $B(x')$ points that could be anywhere. Thus we set $k = B(x')$.

After the initial Metropolis-Hastings update the second part of the random map is a birth-and-death process. With rate K known points uniformly distributed on S are added to D. Whenever a point ξ, say, is added to D it is added to X with probability $\lambda(x, \xi)/K$. This implies that points are born in X with birth rate $\lambda(x, \xi)$. Known points in D die with rate one. When a point in D dies it is also removed from X if it is there. The upper bound on the number

of unknown points k is decreased by one with rate k. Whenever k is decreased by one a uniformly randomly selected point from $X^{(u)}$ is deleted with probability $n(X^{(u)})/k$. Hence, all points in X, unknown as well as known, die with rate one. Together with the birth rate this implies that f is the equilibrium density of the birth-and-death process X, for more detail see e.g. Chapter 2 or (Preston 1977). By construction $n(X^{(u)}) \leq k$, hence when k reaches zero at time T', say, we have $X \subseteq D$.

From time T' until a prespecified time $T > T'$ continue the construction of D and X as above. Additionally, at time T' upper and lower processes are initialised as $U = D$ and $L = \emptyset$ and generate until time T coupled to D as in Chapter 2. If the upper and lower processes have coalesced at time T the random mapping is coalescent and the common state of the upper and lower process is returned as the output of the random mapping. If coalescence is not achieved the state of X is returned which depends on the input state X_0.

The target distribution is invariant for this random mapping as the target distribution is invariant under both the initial Metropolis-Hastings update and the birth-and-death process and because T is fixed and hence independent of the birth-and-death process.

To determine a time T so that $p = \mathbb{P}(\text{random map coalescent}) > 0$ an initial run of the above procedure is executed where the birth-and-death process is continued until the upper and lower processes have coalesced. Then T is set equal to the running time of this initial test run.

Wilson (2000a) argues that the running time of this application of the read-once algorithm is comparable to that of the dominated CFTP algorithm.

Bibliography

Brix, A. & Kendall, W. S. (2002). Simulation of cluster point processes without edge effects, *Adv. Appl. Prob.* **34**: 267–280.

Brooks, S. P., Fan, Y. & Rosenthal, J. S. (2002). Perfect forward simulation via simulated tempering, *Technical Report 2002-05*, University of Cambridge, Statistical Laboraoty. Submitted.

Chan, K. S. & Geyer, C. J. (1994). Discussion of the paper 'markov chains for exploring posterior distributions' by luke tierney, *Ann. Statist.* **22**: 1747–1758.

Cox, D. R. (1955). Some statistical methods connected with series of events, *J. R. Statist. Soc. Ser. B* **17**: 129–164.

Daley, D. J. & Vere-Jones, D. (1988). *An Introduction to Point Processes*, Springer, Berlin.

Dellaportas, P. & Roberts, G. O. (2003). Introduction to MCMC, *in* J. Møller (ed.), *Spatial Statistics and Computational Methods*, Lecture Notes in Statistics 173, Springer-Verlag, New York.

Diggle, P. & Gratton, R. (1984). Monte Carlo methods of inference for implicit statistical model (with discussion), *J. Roy. Statist. Soc. Ser. B* **46**: 193–227.

Diggle, P. J., Besag, J. & Gleaves, J. T. (1976). Statistical analysis of spatial point patterns by means of distance methods, *Biometrics* **32**: 659–667.

Diggle, P. J., Fiksel, T., Grabarnik, P., Ogata, Y., Stoyan, D. & Tanemura, M. (1994). On parameter estimation for pairwise interaction point processes, *Int. Statist. Rev.* **62**: 99–117.

Fill, J. A. (1998). An interruptible algorithm for perfect sampling via Markov chains, *Annals of Applied Probability* **8**: 131–162.

Gelman, A. & Meng, X.-L. (1998). Simulating normalizing constants: from importance sampling to bridge sampling to path sampling, *Statist. Sci.* **13**: 163–185.

Geman, S. & Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**: 721–741.

Geyer, C. J. (1992). Practical Markov Chain Monte Carlo, *Statistical Science* **7**: 473–511.

Geyer, C. J. (1999). Likelihood inference for spatial processes, *in* O. E. Barndorff-Nielsen, W. Kendall & M. van Lieshout (eds), *Stochastic Geometry, Likelihood and Computation*, Chapman & Hall, pp. 79–140.

Geyer, C. J. & Møller, J. (1994). Simulation procedures and likelihood inference for spatial point processes, *Scand. J. Statist.* **21**: 359–373.

Geyer, C. J. & Thompson, E. A. (1995). Annealing Markov chain Monte Carlo with applications to ancestral inference, *J. Amer. Statist. Assoc.* **90**: 909–920.

Gilks, W. R. & Wild, P. (1992). Adaptive rejection sampling for gibbs sampling, *Appl. Statist.* **41**: 337–348.

Green, P. (1995). Reversible jump MCMC computation and Bayesian model determination, *Biometrika* **82**: 711–732.

Häggström, O., van Lieshout, M. N. M. & Møller, J. (1999). Characterization results and Markov chain Monte Carlo algorithms including exact simulation for some spatial point processes, *Bernoulli* **5**: 641–658.

Hastings, W. (1970). Monte carlo sampling methods using markov chains and their applications, *Biometrika* **57**: 97–109.

Kallenberg, O. (1984). An informal guide to the theory of conditioning in point processes, *Int. Statist. Rev.* **52**: 151–164.

Kendall, W. S. (1998). Perfect simulation for the area-interaction point process, *in* L. Accardi & C. Heyde (eds), *Probability Towards 2000*, Springer, New York, pp. 218–234.

Kendall, W. S. & Møller, J. (2000). Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes, *Adv. Appl. Prob.* **32**: 844–865.

Kendall, W. S. & Thönnes, E. (1999). Perfect simulation in stochastic geometry, *Pattern Recognition* **32**: 1569–1586.

Kingman, J. F. C. (1993). *Poisson Processes*, Oxford University Press, Oxford.

Lieshout, M. N. M. van (2000). *Markov Point Processes and their Applications*, Imperial College Press, London.

Marinari, E. & Parisi, G. (1992). Simulated tempering: A new Monte Carlo scheme, *Europhysics letters* **19**: 451–458.

Meng, X.-L. & Wong, W. H. (1996). Simulating ratios of normalizing constants via a simple identity: a theoretical exploration, *Statist. Sinica* **6**.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953). Equations of state calculations by fast computing machines, *J. Chem. Phys.* **21**: 1087–1092.

Meyn, S. P. & Tweedie, R. L. (1993). *Markov Chains and Stochastic Stability*, Springer-Verlag, London.

Mira, A., Møller, J. & Roberts, G. O. (2001). Perfect slice samplers, *J. R. Statist. Soc. Ser. B* **63**: 593–606.

Møller, J. (2003). Shot noise Cox processes. *Adv. Appl. Prob.* **35**: 614–640.

Møller, J. & Nicholls, G. (1999). Perfect simulation for sample-based inference, *Technical report*, R-99-2011, Department of Mathematical Sciences, Aalborg University.

Møller, J. & Waagepetersen, R. P. (2002). Statistical inference for Cox processes, *in* A. B. Lawson & D. G. T. Denison (eds), *Spatial Cluster Modelling*, Chapman & Hall/CRC, Boca Raton, pp. 37–60.

Møller, J. & Waagepetersen, R. P. (2003). *Statistical Inference and Simulation for Spatial Point Processes*, Chapman and Hall/CRC, Boca Raton, Florida.

Neumann, J. von (1951). Various techniques used in connection with random digits, *in* A.S. Householder et al. (ed.), *The Monte Carlo Method*, number 12 in *Nat. Bur. Standards Appl. Math. Ser.*, pp. 36–38.

Norris, J. R. (1997). *Markov Chains*, Cambridge University Press, New York.

Ogata, Y. & Tanemura, M. (1984). Likelihood analysis of spatial point patterns, *J. R. Statist. Soc. Ser. B* **46**: 496–518.

Penttinen, A. (1984). *Modelling Interaction in Spatial Point Patterns: Parameter Estimation by the Maximum Likelihood Method*, Number 7 in Jyväskylä Studies in Computer Science, Economics, and Statistics.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (1992). *Numerical Recipes in C*, 2nd edn, Cambridge University Press, Cambridge.

Preston, C. (1977). Spatial birth-and-death processes, *Bull. Int. Statist. Inst.* **46**: 371–391.

Propp, J. G. & Wilson, D. B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics, *Random Structures and Algorithms* **9**: 223–252.

Ripley, B. (1979). Simulating spatial patterns: dependent samples from a multivariate density. algorithm as 137, *Appl. Statist.* **28**: 109–112.

Ripley, B. D. (1977). Modelling spatial patterns (with discussion), *J. Roy. Statist. Soc. Ser. B* **39**: 172–212.

Ripley, B. D. (1987). *Stochastic Simulation*, Wiley, New York.

Robert, C. P. & Casella, G. (1999). *Monte Carlo Statistical Methods*, Springer, New York.

Roberts, G. O., Gelman, A. & Gilks, W. R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms, *Ann. Appl. Prob.* **7**: 110–120.

Roberts, G. O. & Rosenthal, J. S. (1997). Geometric ergodicity and hybrid Markov chains, *Electron. Comm. Probab.* **2**: 13–25.

Stoyan, D., Kendall, W. S. & Mecke, J. (1995). *Stochastic Geometry and Its Applications*, 2nd edn, Wiley, Chichester.

Strauss, D. J. (1975). A model for clustering, *Biometrika* **62**: 467–475.

Wilson, D. B. (2000a). How to couple from the past using a read-once source of randomness, *Random Structures and Algorithms* **16**: 85–113.

Wilson, D. B. (2000b). Layered multishift coupling for use in perfect sampling algorithms (with a primer on CFTP), *in* N. Madras (ed.), *Monte Carlo Methods*, Vol. 26 of *Fields Institute Communications*, Amer. Stat. Soc., pp. 141–176.

2 Spatial jump processes and perfect simulation

Kasper K. Berthelsen and Jesper Møller

Department of Mathematical Sciences, Aalborg University,
Fredrik Bajers Vej 7G, DK-9220 Aalborg, Denmark

Abstract. Spatial birth-and-death processes, spatial birth-and-catastrophe processes, and more general types of spatial jump processes are studied in detail. Particularly, various kinds of coupling constructions are considered, leading to some known and some new perfect simulation procedures for the equilibrium distributions of different types of spatial jump processes. These equilibrium distributions include many classical Gibbs point process models and a new class of models for spatial point processes introduced in the text.

2.1 Introduction

Many dynamic systems of interacting "objects" evolving in time can be described by a stochastic process $X = \{X_t : t \geq 0\}$, where t denotes time. Often the objects can be viewed as points living in some appropriate space S, so that $X_t \subseteq S$ for all times t. For example, S could be a planar or spatial region; or a space for some geometric objects like line segments or discs. One class of models for such stochastic processes is that of spatial jump processes.

In this contribution we consider the case of finite spatial jump processes, i.e. when X_t is a finite subset of S. Recall that in general a jump process is a continuous-time Markov process $X = \{X_t : t \geq 0\}$ with piecewise constant sample path (Feller 1971). In the special case where each jump consists in adding/deleting exactly one point, we have a spatial birth-and-death process (Preston 1977). This is illustrated in Fig. 2.1 when $S = [0, 1]$ is a unit interval and in Fig. 2.2 when $S = [0, 1]^2$ is a unit square. For a general spatial jump process as considered later in the text, let x and y be finite subsets of S. Then conditionally on the current state $X_t = x$ and the previous history $\{X_s : s < t\}$, the waiting time τ_x to the next jump (i.e. X_s stays in x for $t \leq s < t + \tau_x$, and $X_{t+\tau_x} \neq x$) is exponentially distributed and depends only on x. Furthermore, conditionally on both $X_t = x$, $\{X_s : s < t\}$, and τ_x, the next state $X_{t+\tau_x} = y$ follows some probability measure P_x which only depends

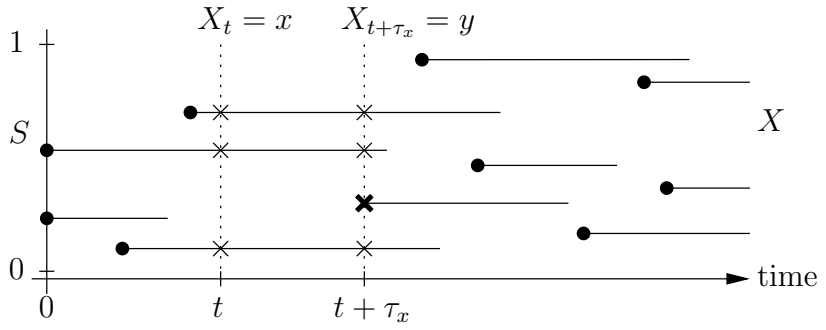

Figure 2.1: Schematic description of a spatial birth-and-death process defined on $S = [0, 1]$. Each horizontal line represents the life span of a point. The point configuration $X_t = x$ at time t is shown (indicated by the crosses at time t). The waiting time τ_x to the next transition (here a birth) is indicated, and the new point configuration $X_{t+\tau_x} = y$ is shown (indicated by the crosses at time $t + \tau_x$, the bold cross being the new born point).

on x so that either $x \subset y$ or $x \supset y$.

Spatial jump processes are interesting for several reasons. They provide a large class of models for spatio-temporal processes, which e.g. may describe many interacting particle systems studied in physics. As we shall demonstrate, they also provide mathematical tractable models for spatio-temporal processes. Specificly, we consider different coupling techniques which allow us to study the ergodic behaviour and characterise the equilibrium distribution for a large class of models in the reversible case as well as in the irreversible case. Furthermore, due to certain thinning techniques and since the sample paths are piecewise constant, spatial jump processes can often easily be simulated on a computer. Particularly, we extend the ideas of Propp & Wilson (1996) and use the thinning techniques for various perfect (or exact) simulation algorithms based on dominated coupling from the past (Kendall 1998, Kendall & Møller 2000). As in Kendall & Møller (2000), we prefer the term "perfect simulation" instead of "exact simulation", since random number generators always have defects and an algorithm may fail to deliver a simulation within practical constraints of time.

Currently perfect simulation is a hot research topic in statistics and applied probability, and as such the perfect simulation part (Sections 2.4 and 2.5.2) is of its own interest. Perfect simulation techniques have proven to be particular useful in statistical physics, spatial statistics, and stochastic geometry. A primer on perfect simulation is provided in Thönnes (2000) and a recent review on perfect simulation in stochastic geometry is given in Møller (2001). Georgii (2000) and Häggström, van Lieshout & Møller (1999) deal with perfect simulation and some aspects of phase transition in the Widom & Rowlinson (1970) model and related models. For a survey on the historic development of perfect simulation and a comprehensive list of references, see `http://www.dbwilson.com/exact`.

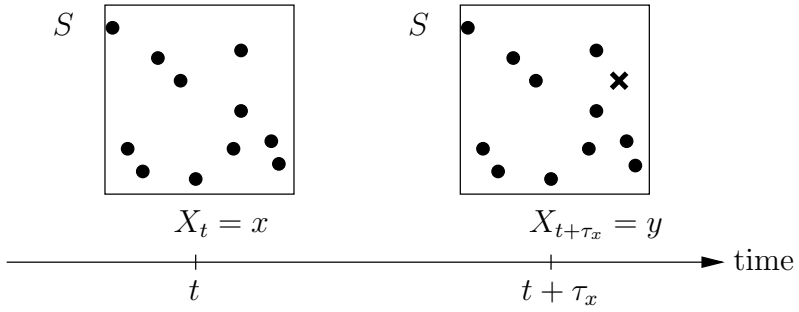

Figure 2.2: Schematic description of a spatial birth-and-death process defined on $S = [0,1]^2$. The point pattern $X_t = x$ at time t is shown, the waiting time τ_x to the next transition (here a birth) is indicated, and the new point configuration $X_{t+\tau_x} = y$ is shown (here the cross is the new born point).

In this contribution we discuss various constructions of spatial jump processes, study the ergodicity properties of these processes, characterise their equilibrium distributions, and show how to start spatial jump processes in equilibrium by the use of perfect simulation techniques. Readers interested in a further discussion of the statistical and computational aspects (using perfect simulations) are referred to Berthelsen & Møller (2001a). Furthermore, many of the ideas in the perfect simulation part of the text apply to Gibbs sampling (also known in statistical physics as the heat-bath algorithm) and other Metropolis-Hastings algorithms for spatial point processes (Häggström et al. 1999, Kendall & Møller 2000, Møller 2001, Møller & Schladitz 1999, Thönnes 1999).

The text is organised as follows. Section 2.2 provides some background material on simple jump processes (i.e. when the position of points is ignored), which becomes useful when we later couple simple jump processes with spatial jump processes. The main part of the text concerns spatial birth-and-death processes. Section 2.3 is concerned with general properties of spatial birth-and-death processes, using certain coupling constructions. A description and comparison of the perfect simulation algorithms in Fernández, Ferrari & Garcia (2000) and Kendall & Møller (2000) using spatial birth-and-death processes is given in Section 2.4. Section 2.5 concerns spatial birth-and-catastrophe processes, i.e. when a jump consists in either adding a new point or deleting all existing points. Section 2.6 comments on more general cases of spatial jump processes. Finally, the main notation is listed and briefly explained in the Appendix.

2.2 Simple jump processes

This section is a short diversion into simple jump processes as they will eventually control everything which goes on in this text. See for example Asmussen (1987) and Norris (1997) for more details.

A simple jump process $N = \{N_t : t \geq 0\}$ is a continuous time Markov process with state space $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$ and piecewise constant right-continuous paths. It can formally be constructed as follows. Let $J_1 < J_2 < J_3 \ldots$ be the times at which N makes a jump, and set $J_0 = 0$. Further, for all $n \in \mathbb{N} = \{1, 2, \ldots\}$, let

$$H_n = \begin{cases} J_n - J_{n-1} & \text{if } J_n < \infty \\ \infty & \text{otherwise} \end{cases}$$

be the holding times, and set $M_n = N_{J_n}$. The discrete time Markov chain $M = \{M_n : n \in \mathbb{N}_0\}$ is called the jump chain (or the embedded Markov chain). For any $n \in \mathbb{N}$ and any $i \in \mathbb{N}_0$, the conditional distribution of H_n given $H_1, \ldots, H_{n-1}, M_0, \ldots, M_{n-1} = i$ is $\text{Exp}(q_i)$, the exponential distribution with mean $1/q_i$, where $q_i > 0$ is a given parameter. Furthermore, for any $j \in \mathbb{N}_0 \backslash \{i\}$,

$$\mathbb{P}(M_n = j | H_1, \ldots, H_n, M_0, \ldots, M_{n-1} = i) = q_{ij}/q_i,$$

where the $q_{ij} \geq 0$ are given parameters so that $q_i = \sum_{j \neq i} q_{ij}$. The matrix $Q = \{q_{ij} : i, j \in \mathbb{N}_0\}$ with $q_{ii} = -q_i$ is called the generator of N.

In other words, conditionally on $N_t = i$, the waiting time τ_i to the next jump is independent of the previous history $\{N_s : s < t\}$, and $\tau_i \sim \text{Exp}(q_i)$. Moreover, conditionally on both $N_t = i$, $\{N_s : s < t + \tau_i\}$, and τ_i, we have that $N_{t+\tau_i}$ jumps to j with probability q_{ij}/q_i for $j \neq i$.

In the sequel we impose the following technical conditions, which are commented below.

Conditions:
(i) $\pi = \{\pi_i : i \in \mathbb{N}_0\}$ is a given probability density function where the support $I = \{i : \pi_i > 0\}$ is given by either $I = \mathbb{N}_0$ or $I = \{0, \ldots, l\}$ for some $l \in \mathbb{N}_0$.
(ii) Q is irreducible, i.e. for all distinct $i, j \in I$ there exist $i_0, i_1, \ldots i_n \in I$ such that $i_0 = i$, $i_n = j$, and $q_{i_0 i_1} \ldots q_{i_{n-1} i_n} > 0$.
(iii) π is invariant for Q, i.e. $\pi Q = 0$.
(iv) Q is non-explosive, i.e. $\mathbb{P}(\sum_n H_n < \infty | N_0 = i) = 0$ for all $i \in I$; see Fig. 2.3.

Remarks:
(a) The density π will play a key role in the sequel. That π is invariant for Q means that if N_s follows π, then N_t follows π for all $t \geq s$. Condition (i) is a kind of hereditary condition on π. It is equivalent to assuming that the support is of the form $I = \{0, 1, \ldots, l\}$ or $I = \mathbb{N}_0$.
(b) Conditions (ii) and (iii) imply that π is the unique invariant distribution for Q.

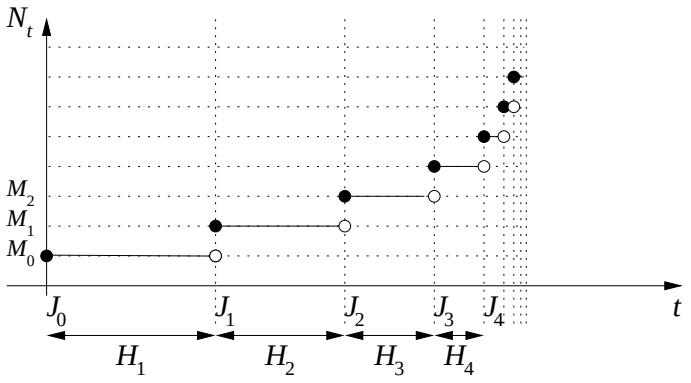

Figure 2.3: Schematic description of a simple jump process which does not satisfy the non-explosive condition.

(c) Sometimes we assume that Q and π are in detailed balance, i.e.

$$\pi_i q_{ij} = \pi_j q_{ji} \text{ for all } i, j \in \mathbb{N}_0. \tag{2.1}$$

(d) Detailed balance implies that π is invariant for Q. Combining (ii), (iv), and (2.1) we obtain reversibility of N.
(e) Condition (iv) holds if and only if the only bounded solution to $Qk = k$ for column vectors $k = (k_0, k_1, \ldots)^{\mathrm{T}}$ is $k = (0, 0, \ldots)^{\mathrm{T}}$ (see e.g. Asmussen (1987)). A sufficient condition for this is $\sup_{i \in I} q_i < \infty$.
(f) Conditions (ii) and (iv) imply positive recurrence, i.e. for any $i \in I$, the time between two consecutive occurrences of i in N has finite mean.
(g) Conditions (i)–(iv) imply convergence towards π, see e.g. Theorem 3.6.2 in Norris (1997) or Theorem II.4.6 in Asmussen (1987).
(h) For later purposes we notice that the jump chain has invariant density $\pi'_i \propto q_i \pi_i$, and it is positive recurrent (by Theorems 3.4.1 and 3.5.3 in Norris (1997), the jump chain is recurrent; and recurrence implies positive recurrence as the support I is countable). Furthermore, let L be the number of jumps in a cycle of the jump chain, i.e. the number of jumps between two successive zeros in the jump chain. Then by Kac's theorem (Meyn & Tweedie 1994),

$$\mathbb{E}L = 1/\pi'_0. \tag{2.2}$$

2.3 Spatial birth-and-death processes

This section considers finite spatial birth-and-death processes defined on some rather arbitrary space S. Section 2.3.1 concerns the simplest case where S

consists of one element only; this case corresponds to a simple jump process with jumps of the type $i \to i \pm 1$. Section 2.3.2 introduces a general notation for both spatial birth-and-death processes and the spatial jump processes considered later in this text. The general case of a spatial birth-and-death process is described in Section 2.3.3, using a coupling to a dominating simple birth-and-death process. Section 2.3.4 discusses a detailed balance condition and an extension of the concept of local stability (Ruelle 1969). Section 2.3.5 considers how to construct in an easy way a spatial birth-and-death process by thinning from a simpler dominating spatial birth-and-death process, extending ideas in Kendall (1998) and Kendall & Møller (2000). This coupling construction will also be used in Section 2.4 for perfect simulation purposes.

2.3.1 The simple case

A simple jump process with generator Q is a simple birth-and-death process if $q_{ij} = 0$ whenever $|i - j| > 1$. For convenience set $b_i = q_{i,i+1}$ and $d_{i+1} = q_{i+1,i}$ for $i \in \mathbb{N}_0$. Throughout this section we assume that

$$\pi_i b_i = \pi_{i+1} d_{i+1} > 0 \quad \text{whenever } \pi_{i+1} > 0. \tag{2.3}$$

This assumption is equivalent to the conditions (i)–(iii) and the detailed balance condition (2.1) in Section 2.2. Regarding the final condition (iv) in Section 2.2, by Reuter & Ledermann (1953), a simple birth-and-death process is non-explosive if

$$\sum_{i \in I} \frac{1}{b_i} = \infty. \tag{2.4}$$

In the sequel we often refer to the following Examples 1–3.

Example 1: Consider a standard immigration birth-and-death process, i.e. $b_i \equiv \beta > 0$ and $d_{i+1} = i + 1$ for $i \in \mathbb{N}_0$. Clearly, (2.3) and (2.4) are satisfied, and

$$\pi_i = e^{-\beta} \beta^i / i!, \tag{2.5}$$

i.e. π specifies a Poisson distribution with mean β. By (h) in Section 2.2, $\pi'_{i+1} \propto \pi_{i+1} + \pi_i$. Hence by (2.2), $\mathbb{E}L = 2e^{\beta}$.

Example 2: For $i \in \mathbb{N}_0$, let $b_i = p$ and $d_{i+1} = 1 - p$, where $0 < p < \frac{1}{2}$. Then (2.3) and (2.4) are satisfied, and

$$\pi_i = (1 - q) q^i \tag{2.6}$$

specifies a geometric distribution with parameter $q = p/(1 - p)$. The corresponding jump chain is a random walk with reflecting barrier at 0. Using (h) and (2.2), we obtain that $\mathbb{E}L = 2q/(1 - q)$.

Example 3: In this example we start by specifying π and then determine the birth and death rates such that (2.3) and (2.4) hold. Assume that

$$\pi_i = \frac{\Gamma(i+\alpha)}{\Gamma(\alpha)i!}(1-q)^\alpha q^i \tag{2.7}$$

is negative binomial, where $\alpha > 0$ and $0 < q < 1$. Letting $q = p/(1-p)$, one solution to (2.3) is an extension of Example 2 such that

$$b_i = p(i+\alpha)/(i+1) \quad \text{and} \quad d_{i+1} = 1-p.$$

Then (2.4) is seen to hold.

2.3.2 Notation for spatial jump processes

We now consider some general notation which applies not only to spatial birth-and-death processes but all spatial jump processes considered in this text. We avoid topological and measure theoretical details; such details can be found in e.g. Daley & Vere-Jones (1988), van Lieshout (2000) and Berthelsen & Møller (2002).

In the spatial case of jump processes $X = \{X_t : t \geq 0\}$, we consider this as a stochastic processes defined on a metric space S with metric $d(\cdot,\cdot)$; in many of our examples, $S = [0,1]^2$ is a unit square and $d(\xi,\eta) = \| \xi - \eta \|$ is the Euclidian distance. For simplicity we consider X_t to be a finite subset of S, though everything in the sequel easily extend to the case where X_t is allowed to have multiple points. For any finite $x \subseteq S$, we let $n(x)$ denote the number of points in x. Then the state space of X_t is $\Omega = \bigcup_{i=0}^{\infty} \Omega_i$, where $\Omega_i = \{x \subseteq S : n(x) = i\}$ is the set of finite point configurations of cardinality i; note that $\Omega_0 = \{\emptyset\}$ where $\emptyset$ denotes the empty point configuration. For $x \in \Omega$ and $\xi \in S$, we write $x \backslash \xi$ for $x \backslash \{\xi\}$, and $x \cup \xi$ for $x \cup \{\xi\}$.

Finally, we let λ denote an arbitrary diffuse probability measure on S (here "diffuse" means that $\lambda(\{\xi\}) = 0$ for all $\xi \in S$). For example, if $S \subset \mathbb{R}^d$ has finite volume, λ is typically taken to be the uniform distribution on S.

2.3.3 Description of spatial birth-and-death processes

Now, consider a spatial birth-and-death process X. For the following description of X the reader may find it illuminating to consider Fig. 2.1 and Fig. 2.2 in Section 2.1.

We specify the distribution of X by two functions $\beta, \delta : \Omega \times S \to [0,\infty)$. For any $x \in \Omega$, $\beta(x,\xi)$ is the birth rate at which a point $\xi \in S$ is added to x, and if $x \neq \emptyset$, $\delta(x \backslash \eta, \eta)$ is the death rate at which a point $\eta \in x$ is deleted from x. Roughly speaking, for an infinitesimal time interval $[t, t+dt]$ and an

infinitesimal ball $d\xi$ with centre ξ,

$$\begin{aligned}\mathbb{P}(X_{t+dt} \setminus x \text{ is a point in } d\xi \,|X_t = x) &\approx \beta(x,\xi)\lambda(d\xi)dt,\\ \mathbb{P}(X_{t+dt} = x\backslash\eta \,|X_t = x) &\approx \delta(x\backslash\eta,\eta)dt,\\ \mathbb{P}(\text{more than one jump in } [t,t+dt]|X_t = x) &\approx 0.\end{aligned}$$

More precisely, define in a similar way as in Section 2.2, the jump times $J_1 < J_2 < \ldots$, the holding times $H_1, H_2, \ldots$, and the jump chain $Y = \{Y_n : n \in \mathbb{N}_0\}$ for X, where $Y_0 = X_0$. Further, for $x \in \Omega$, let $B(x) = \int_S \beta(x,\xi)\lambda(d\xi)$ and $D(x) = \sum_{\eta\in x} \delta(x\backslash\eta,\eta)$ be the total birth and death rates, assuming $B(x) < \infty$ and setting $D(\emptyset) = 0$. Then the conditional distribution of H_n given $H_1,\ldots,H_{n-1}$, $Y_0,\ldots,Y_{n-1} = x$ is $\mathrm{Exp}\,(B(x)+D(x))$. If we also condition on H_n, the probability for a birth at time J_n is $B(x)/(B(x)+D(x))$. If we further condition on that a birth happens at time J_n, then $Y_n = x\cup\xi$, where ξ has density $\beta(x,\cdot)/B(x)$ with respect to λ. If we instead condition on that a death happens at time J_n, then $Y_n = x\backslash\eta$, where $\eta \in x$ is selected with probability $\delta(x\backslash\eta,\eta)/D(x)$.

It remains to clarify if such a spatial birth-and-death process is non-explosive. Given a generator Q for a non-explosive simple birth-and-death process N as in Section 2.3.1, we assume henceforth that

$$\beta(x,\xi) = b_{n(x)}\bar{b}(x,\xi) \quad \text{and} \quad \delta(x,\xi) = d_{n(x)+1}\bar{d}(x,\xi)/(n(x)+1) \tag{2.8}$$

where the functions $\bar{b}$ and $\bar{d}$ satisfy the following constraints:

$$0 \le \bar{b} \le 1 \quad \text{and} \quad \bar{d} \ge 1. \tag{2.9}$$

By (2.8) and (2.9),

$$B(x) \le b_{n(x)} \quad \text{and} \quad D(x) \ge d_{n(x)}. \tag{2.10}$$

It is then possible to couple N and X so that

$$N_t \ge n(X_t) \quad \text{for all } t \ge 0 \tag{2.11}$$

if $N_0 \ge n(X_0)$. Briefly, (X,N) is a jump process where X and N evolve independently of each other as long as $N_t > n(X_t)$, while if $N_t = n(x) = n$ and $X_t = x$, (X_t, N_t) jumps to (x',n') with a rate specified as follows:

$$\begin{aligned}&\beta(x,\xi) \ \text{ for } (x',n') = (x\cup\xi, n+1),\\ &b_n - B(x) \ \text{ for } (x',n') = (x, n+1),\\ &d_n\delta(x\backslash\eta,\eta)/D(x) \ \text{ for } (x',n') = (x\backslash\eta, n-1),\\ &(D(x)-d_n)\delta(x\backslash\eta,\eta)/D(x) \ \text{ for } (x',n') = (x\backslash\eta, n).\end{aligned}$$

By (2.10) and (2.11), since Q is non-explosive, it is intuitively clear that X is non-explosive. A formal proof can be found in Preston (1977).

Combining (f) in Section 2.2, (2.11), and the renewal theorem (see e.g. Theorem V.1.2 in Asmussen (1987)), we obtain the following results. Recall that $\emptyset$ denotes the empty point configuration. Write X_{t-} for the state of X just before time t, let $\mathcal{T} = \inf\{t > 0 : X_{t-} \neq \emptyset, X_t = \emptyset\}$ be the return time to $\emptyset$ when $X_0 = \emptyset$, and set $\mu = \mathbb{E}(\mathcal{T}|X_0 = \emptyset)$. Then $\emptyset$ is an ergodic atom, i.e. $\mu < \infty$. Hence X jumps to $\emptyset$ infinite often, and at each such jump time X regenerates (i.e. X at such a time is independent of its past history). As $t \to \infty$, X_t converges in distribution towards an equilibrium distribution κ, say, independent of the initial state of X_0, i.e. for all bounded functions $f : \Omega \to \mathbb{R}$ which are continuous almost everywhere with respect to κ ,

$$\mathbb{E}[f(X_t)|X_0] \to \int f(x)\kappa(dx) \quad \text{as } t \to \infty. \tag{2.12}$$

Further,

$$\mu \int f(x)\kappa(dx) = \mathbb{E}\left[\int_0^{\mathcal{T}} f(X_t)dt \middle| X_0 = \emptyset\right], \tag{2.13}$$

and the support of κ is given by

$$\begin{aligned}\operatorname{supp}(\kappa) = \Omega_0 \cup \big\{\{x_1, \ldots, x_n\} \in \Omega : n \in \mathbb{N} \text{ and} \\ \beta(\{x_1, \ldots, x_{i-1}\}, x_i) > 0 \text{ for } i = 1, \ldots, n\big\}.\end{aligned} \tag{2.14}$$

Finally, we often have geometrically fast convergence towards κ, cf. Møller (1989). For instance, this is the case when b_i and d_i are specified as in Examples 1–3.

2.3.4 Detailed balance and local stability conditions

Let the situation be as in Section 2.3.3. In this section we try to get hold on the equilibrium distribution κ in (2.13). Particularly, we discuss a detailed balance condition which characterises κ in terms of an (unnormalised) density ϕ. Furthermore, we motivate an extension of the important concept of local stability (Ruelle 1969, Geyer 1999, Kendall & Møller 2000).

In order to specify the invariant distribution κ, we assume that ϕ is an unnormalised density with respect to the probability measure on Ω given by

$$\nu = \sum_{i=0}^{\infty} \pi_i \lambda_i. \tag{2.15}$$

Here π is given by (2.3), λ_0 is the probability measure concentrated at the empty point configuration, and λ_i ($i \geq 1$) denotes the distribution for a binomial point process given by i independent points in S with common distribution λ. Note that for a point process $Z \sim \phi$, conditionally on $n(Z) = i$, Z has unnormalised density ϕ with respect to λ_i, i.e. no matter the choice of π. In most applications π is a Poisson distribution with mean β. Then ν is simply a Poisson point

process with intensity measure $\beta\lambda$. Moreover, if λ is the uniform distribution on a bounded set $S \subset \mathbb{R}^d$, then $-\ln\beta$ is the so-called chemical activity for a point process $Z \sim \phi$.

Now, as in Ripley (1977), we obtain that the detailed balance condition,

$$\bar{b}(x,\xi)\phi(x) = \bar{d}(x,\xi)\phi(x \cup \xi) > 0 \quad \text{whenever } \phi(x \cup \xi) > 0 \tag{2.16}$$

ensures that κ has a density proportional to ϕ, and X is time reversible with respect to κ (Ripley (1977) assumes π to be Poisson, but we do not need to make this restriction). Note that (2.9) and (2.16) always imply that ϕ is non-increasing,

$$\phi(x \cup \xi) \leq \phi(x) \tag{2.17}$$

which in turn implies integrability of ϕ with respect to ν no matter the choice of π. We refer to (2.17) as local stability of κ with respect to ν for the following reasons.

Geyer (1999) and Kendall & Møller (2000) consider the common case (2.5) where π is Poisson and

$$\phi(x \cup \xi) \leq K\phi(x), \quad \bar{d} \equiv 1, \quad \bar{b}(x,\xi) = \begin{cases} \phi(x \cup \xi)/\phi(x) & \text{if } \phi(x) \neq 0 \\ 0 & \text{otherwise} \end{cases} \tag{2.18}$$

for some constant K (independent of x and ξ). Then points are dying with uniform rate 1, and $\bar{b}$ is the Papangelou conditional intensity (Kallenberg 1984, Daley & Vere-Jones 1988). Note that (2.18) implies (2.16), and the existence of the uniform upper bound K on $\bar{b}$ is local stability in the sense of Ruelle (1969). As we can replace β in (2.5) by $\beta' = \beta K$, and ϕ in (2.18) by $\phi'(x) = \phi(x)/K^{n(x)}$, we can without loss of generality assume $K = 1$. Then (2.9) and (2.17) are satisfied.

On the other hand, if π is not Poisson, then in general we do not have Ruelle local stability. For instance, let π be given by (2.6), and let $\tilde{\nu} = \sum_i \tilde{\pi}_i \lambda_i$ where $\tilde{\pi}_i = e^{-1}/i!$ is Poisson. Then $\tilde{\phi}(x) = \phi(x)q^{n(x)}n(x)!$ is the corresponding unnormalised density with respect to $\tilde{\nu}$, but (2.17) does not ensure that

$$\tilde{\phi}(x \cup \xi)/\tilde{\phi}(x) = (\phi(x \cup \xi)/\phi(x))q(n(x)+1)$$

is bounded by a constant K.

We now consider two examples where (2.17) is satisfied. The first example is frequently used later in the text, and it shows the effect of the specification of π. The other demonstrates that the specification of π and ϕ sometimes requires a little thought.

Example 4: Suppose that

$$\phi(x) = \gamma^{s_R(x)} \tag{2.19}$$

where $s_R(x) = \sum_{\{\xi,\eta\} \subseteq x} \mathbf{1}[d(\xi,\eta) \leq R]$, and $0 \leq \gamma \leq 1$ and $R \geq 0$ are given parameters. In the case where π is Poisson, we obtain a usual Strauss process (Strauss 1975, Kelly & Ripley 1976); if further $\gamma = 0$, we have a Gibbs hard

core process (taking $0^0 = 1$ in (2.19); see Example 7 below). Clearly, ϕ is non-increasing.

Simulated realisations of different Strauss processes are shown in Fig. 2.4. For this the perfect simulation algorithm in Section 2.4.2 has been used. Due to the thinning procedure in that algorithm (as illustrated later in Fig. 2.6), the first point pattern in Fig. 2.4 contains the two others.

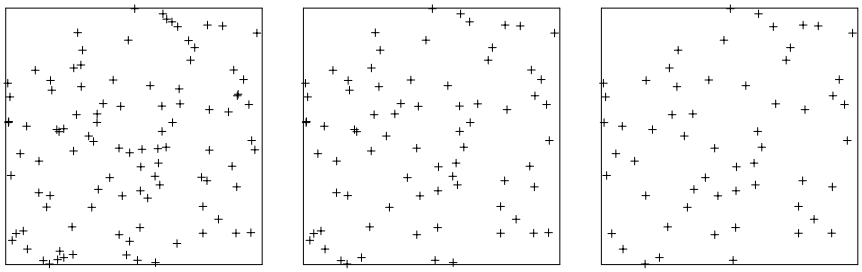

Figure 2.4: Simulation of a usual Strauss process ($\pi_i \propto \beta^i/i!$) on $S = [0,1]^2$, when $\beta = 100$, $R = 0.05$, and $\gamma = 1, 0.5, 0$ (from left to right).

Fig. 2.5 shows the mean number of points for different values of γ when π is either Poisson or geometric; the means are estimated by Monte Carlo methods using the perfect simulation algorithms in Sections 2.4.2 and 2.5.2. In the geometric case, for γ close to 1, the mean is very sensitive to changes in γ, while for γ not close to 1, the number of points is very small.

Example 5: Assume that $S \subset \mathbb{R}^k$ has Lebesgue measure 1, λ is the uniform distribution on S, and $\pi_i \propto \beta^i/i!$ is Poisson. Consider an area interaction point process with unnormalised density

$$\phi(x) = \gamma^{-\lambda(U_x)}$$

with respect to ν (Widom & Rowlinson 1970, Baddeley & van Lieshout 1995, Häggström et al. 1999, Møller 2001). Here $U_x = \cup_{\xi \in x} \text{ball}(\xi, R)$ where $\text{ball}(\xi, R)$ denotes the closed ball centred in ξ with radius R, and $R > 0$ and $\gamma \geq 0$ are given parameters. The area interaction process and its generalisations are relevant in physics, see (Mecke 2000) and the references therein.

The area interaction process is said to be attractive for $\gamma > 1$ and repulsive for $\gamma < 1$, as

$$\phi(x \cup \xi)/\phi(x) = \gamma^{-\lambda(U_{x \cup \xi} \setminus U_x)} \tag{2.20}$$

is non-decreasing ($\gamma > 1$) or non-increasing ($\gamma < 1$) in x. It follows from (2.20) that (2.17) holds in the attractive case but not in the repulsive case. In the latter case we therefore redefine ν by taking $\pi_i \propto \left(\beta\gamma^{-B}\right)^i/i!$ where B is the volume of a ball in $\mathbb{R}^k$ with radius R, and redefine ϕ by

$$\phi(x) = \gamma^{n(x)B - \lambda(U_x)}.$$

Then (2.17) is satisfied.

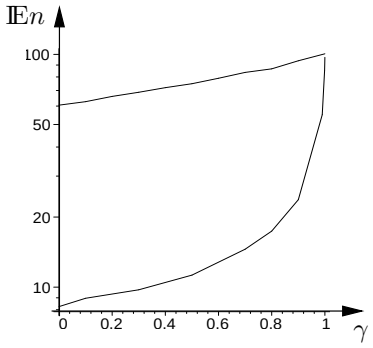

Figure 2.5: Mean number of points versus γ in a usual Strauss process ($\pi_i \propto \beta^i/i!$; upper curve) and in a "geometric-Strauss" process ($\pi_i \propto q^i$; lower curve). In both models, $R = 0.05$ and π has mean 100 (i.e. $\beta = 100$ and $q = 100/101$). Each mean number of points is estimated from 500 i.i.d. samples.

2.3.5 Coupling construction

Consider again the situation in Section 2.3.3. This section presents a coupling constructing where X is obtained by thinning from a dominating spatial birth-and-death process D, which is easily constructed. The coupling construction extends that of Kendall (1998) and Kendall & Møller (2000), and it is advantageous for simulation of X as the computation of the integral $B(x) = \int_S \beta(x,\xi)\lambda(d\xi)$ is not needed. Furthermore, the coupling construction becomes later useful for perfect simulation.

As will be clarified below, we need to assume that

$$b_{i-1} \le b_i \quad \text{and} \quad d_i/i \ge d_{i+1}/(i+1) \quad \text{for } i \in I\backslash\{0\}. \tag{2.21}$$

For instance, (2.21) is satisfied in Examples 1–3.

The dominating process D has birth and death rates

$$\beta^D(x,\xi) = b_{n(x)} \quad \text{and} \quad \delta^D(x,\xi) = d_{n(x)+1}/(n(x)+1).$$

This choice corresponds to a maximal birth rate and a minimal death rate in (2.8), cf. (2.9). By (2.16), D has invariant distribution ν. As the total birth and death rates satisfy

$$B^D(x) = b_{n(x)} \quad \text{and} \quad D^D(x) = d_{n(x)},$$

we can take $N = n(D)$. The process $D_t,\ t \ge 0$, can easily be generated: The simple birth-and-death process N is straightforwardly generated, and $D_t,\ t \ge 0$, is constructed forwards in time by

$$\begin{aligned} N_t = N_{t-} + 1 \quad &\Rightarrow \quad D_t = D_{t-} \cup \xi_t, \\ N_t = N_{t-} - 1 \quad &\Rightarrow \quad D_t = D_{t-}\backslash\eta_t, \end{aligned}$$

where $\xi_t \sim \lambda$ and η_t is a uniformly selected point from D_{t-} (and where ξ_t and η_t are independent of the history of D before time t).

We wish to couple X to D so that D dominates X in the sense that

$$X_t \subseteq D_t \quad \text{for all } t \geq 0. \tag{2.22}$$

For simplicity, as in Kendall (1998) and Kendall & Møller (2000), we consider the case where $d_i \propto i$. This is fulfilled in Example 1. Further we assume that $\bar{d}$ and $\bar{b}$ are specified by (2.18) where ϕ satisfies (2.17). Then X_t can be constructed iteratively for $t \geq 0$ as follows. Each time $N_t = N_{t-} + 1$ we associate to the birth in D at time t a mark $R_t \sim \text{Uniform}[0,1]$ which is independent of $\{D_s : 0 \leq s \leq t\}$ and previous marks $R_s,\ 0 \leq s < t$. Initially let $X_0 \subseteq D_0$ and then

$$N_t = N_{t-} \quad \Rightarrow \quad X_t = X_{t-} \tag{2.23}$$

$$N_t = N_{t-} + 1 \quad \Rightarrow \quad X_t = \begin{cases} X_{t-} \cup \xi_t & \text{if } R_t \leq \bar{b}(X_{t-}, \xi_t) b_{n(X_{t-})} / b_{N_{t-}} \\ X_{t-} & \text{otherwise} \end{cases} \tag{2.24}$$

$$N_t = N_{t-} - 1 \quad \Rightarrow \quad X_t = X_{t-} \backslash \eta_t. \tag{2.25}$$

Clearly this construction satisfies (2.22) as illustrated in Fig. 2.6. It is straightforwardly verified that the birth and death rates of X are given by (2.8). Note that (2.22) and (2.24) imply the need of the first restriction in (2.21).

Example 3 (and hence also Example 2) does not satisfy $d_i \propto i$, but it is still to some extend possible to use the coupling construction above, letting now $d_i = (1-p)i$ and $b_i = (i+\alpha)p$. Then the detailed balance condition (2.3) is fulfilled, and Q is non-explosive as (2.4) is satisfied.

Now, consider the situation where we neither assume $d_i \propto i$ nor that $\bar{b}$ and $\bar{d}$ are specified by (2.18). In general we have that (2.22) together with (2.25) imply a lower bound on the death rate

$$\delta(x \backslash \eta, \eta) = d_{n(x)} \bar{d}(x \backslash \eta, \eta) / n(x) \geq d_{n(y)} / n(y) \quad \text{for } \eta \in x, y \in \Omega, x \subseteq y, x \neq \emptyset.$$

This explains why the second restriction in (2.21) is needed. By (2.8), for each $\eta \in X_{t-}$, we need a rate

$$d_{n(X_{t-})} \bar{d}(X_{t-} \backslash \eta, \eta) / n(X_{t-}) - d_{n(D_{t-})} / n(D_{t-}) \tag{2.26}$$

for deleting η from X but not from D. In the coupling construction below we show how this is possible. However, readers who prefer to skip these technical details should move on to Section 2.4 where only the construction in (2.23)–(2.25) is used.

The process X_t is iteratively generated forwards in time $t \geq 0$ as follows. Initially let $X_0 \subseteq D_0$, and let $X_t = X_{t-}$ whenever $N_t = N_{t-}$. Consider first the case $t = 0$. Set $x = X_t$, $i = n(x)$, and $j = n(D_t)$. Let $\tau = \min\{s > t : N_{s-} \neq N_s\}$ be the first jump time in N after time t. In accordance with (2.26), for each $\eta \in x$, define

$$a_\eta = d_i \bar{d}(x \backslash \eta, \eta)/i - d_j / j \quad \text{and} \quad A(x) = \sum_{\eta \in x} a_\eta, \tag{2.27}$$

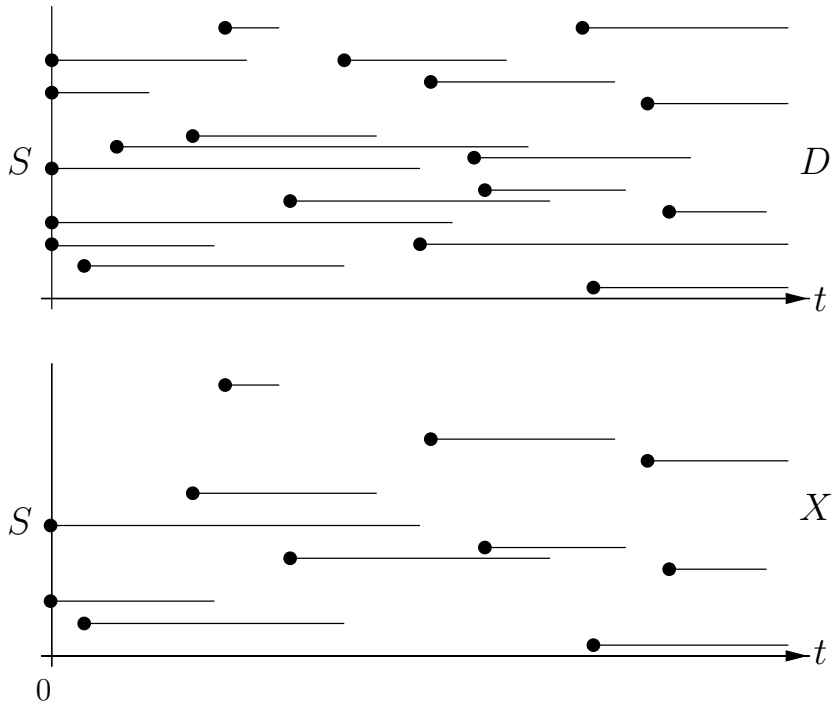

Figure 2.6: Illustration of the coupling construction when $\bar{d} \equiv 1$, $X_0 \subseteq D_0$ and $S = [0,1]$. Top: the dominating process D. Bottom: the target process X obtained by thinning from D as described in (2.24).

setting $A(x) = 0$ if $x = \emptyset$. Then generate $\tau' \sim \text{Exp}(A(x))$ (independently of τ and the history of D and X before and including time t), setting $\tau' = \infty$ if $A(x) = 0$. If $\tau' < \tau - t$, then increase t by τ' and set $X_t = x \backslash \eta$, where $\eta \in x$ is chosen with probability $a_\eta / A(x)$ (independently of τ, τ', and the history of D and X before time t). Else set $t = \tau$ and let

$$N_t = N_{t-} + 1 \quad \Rightarrow \quad X_t = \begin{cases} x \cup \xi_t & \text{if } R_t \leq \bar{b}(x, \xi_t) b_i / b_j \\ x & \text{otherwise} \end{cases} \tag{2.28}$$

$$N_t = N_{t-} - 1 \quad \Rightarrow \quad X_t = x \backslash \eta_t, \tag{2.29}$$

where in (2.28) we associate to the birth in D a mark $R_t \sim \text{Uniform}[0,1]$ (independent of τ, τ', the history of D before and including time t, and the history of X before time t). Repeating this procedure it is straightforwardly verified that (2.22) is satisfied and the birth and death rates of X are given by (2.8).

2.4 Perfect simulation

In this section we discuss perfect simulation procedures based on coupling from the past (Propp & Wilson 1996) for a point process with a distribution κ with respect to ν in (2.15). We assume that κ is the equilibrium distribution of a spatial birth-and-death process satisfying the conditions in Sections 2.3.3 and 2.3.5. We use the coupling construction in Section 2.3.5 in two ways, extending the methods of Kendall & Møller (2000) in Sections 2.4.1 and 2.4.2, and of Fernández et al. (2000) in Section 2.4.3. Finally, in Section 2.4.3 we compare the methods and report on some empirical findings.

Throughout this section we assume for simplicity that $d_i \propto i$, and $\bar{d}$ and $\bar{b}$ are specified by (2.18) where ϕ satisfies (2.17). However, it is possible to extend our perfect samplers to the case of the more complicated coupling construction described at the end of Section 2.3.5.

2.4.1 Dominated coupling from the past

In this section we describe the simplest version of dominated coupling from the past (dominated CFTP), also called horizontal CFTP, cf. the survey in Møller (2001). For further details on dominated CFTP in a general setting, see Kendall & Møller (2000). Refined versions of dominated CFTP are given in Sections 2.4.2 and 2.4.3.

The basic idea in dominated CFTP is as follows. Start the dominating process D from Section 2.3.5 in equilibrium at time 0, and extend it both forwards and backwards in time. This is generally easily done, since D has equilibrium distribution ν and D is time reversible. Further, for all forwards births $D_t \backslash D_{t-} \neq \emptyset$ with $-\infty < t < \infty$, generate independent marks $R_t \sim \text{Uniform}[0,1]$, where the R_t's are independent of $\{D_t : -\infty < t < \infty\}$. Let $\tau_i,\ i \in \mathbb{N}_0$, denote the times D enters $\emptyset$, i.e. $D_{\tau_i -} \neq \emptyset$ and $D_{\tau_i} = \emptyset$, and assume that $\tau_i < \tau_{i+1}$ for all $i \in \mathbb{N}_0$. For each $i \in \mathbb{N}_0$, construct a target process X_t on the time interval $[\tau_i, \tau_{i+1}]$ in exactly the same way as in Section 2.3.5 except that we are now starting with $X_{\tau_i} = \emptyset$. Then (D, X) regenerates each time D enters $\emptyset$, $X_t \subseteq D_t$ for all $-\infty < t < \infty$, and X is a spatial birth-and-death process with equilibrium distribution κ, where κ has unnormalised density ϕ. Because of time-stationarity of (D, X), for any two fixed times, e.g. 0 and -1, X_0 and X_{-1} follow the same distribution, which is then the invariant distribution κ. Hence $X_0 \sim \kappa$.

Let $-T_\emptyset \leq 0$ denote the first time that D enters $\emptyset$ when D is generated backwards from time 0. As $\emptyset$ is an ergodic atom, $T_\emptyset$ is finite almost surely. Note that for the generation of X_0, it suffices to consider the jump chains of D and X on the time interval $[-T_\emptyset, 0]$. Let $T'_\emptyset$ be the number of jumps D makes on $[-T_\emptyset, 0]$. As illustrated in Fig. 2.7, for the perfect simulation procedure described above, we need only to

(I) generate $D_0 \sim \nu$;

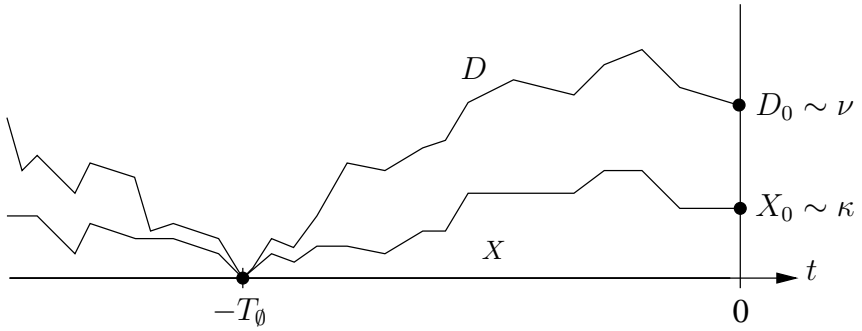

Figure 2.7: Illustration of the idea behind dominated CFTP (coupling from the past).

(II) generate the jump chain of D (together with the marks for forwards births), for $T'_\emptyset$ steps backwards in time from time 0;

(III) construct the jump chain of $\{X_t : -T_\emptyset \leq t \leq 0\}$ forwards in time;

(IV) return $X_0 \sim \kappa$.

However, for many applications $T'_\emptyset$ will be infeasibly large as shown in the following example.

Example 6: Consider the jump chain $\{\ldots, M_{-2}, M_{-1}, M_1, M_2, \ldots\}$ from Section 2.2 when this is in equilibrium and extended to all times $n \in \mathbb{Z}\backslash\{0\}$. Set $M_0 = M_{-1}$, $L_0 = T'_\emptyset + T''_\emptyset$ if $M_0 \neq 0$, and $L_0 = \inf\{n \in \mathbb{N} : M_n = 0\}$ if $M_0 = 0$, where $T'_\emptyset = \inf\{n \in \mathbb{N} : M_{-n} = 0\} - 1$ and $T''_\emptyset = \inf\{n \in \mathbb{N}_0 : M_n = 0\}$. By reversibility, $T'_\emptyset$ and $T''_\emptyset$ are identically distributed. Further, conditionally on $M_0 = 0$, we have that $L, T'_\emptyset, T''_\emptyset$ are identically distributed. Furthermore, by time-stationarity, $\mathbb{P}(L_0 = l) = \mathbb{P}(L = l)l/\mathbb{E}L$, hence $\mathbb{E}L_0 \geq \mathbb{E}L$, i.e. $\mathbb{E}T'_\emptyset \geq \mathbb{E}L/2$. Combining these properties with (2.2), it is straightforwardly derived that

$$\mathbb{E}T'_\emptyset \geq (1 + 1/\pi'_0)/2. \tag{2.30}$$

In the examples below we just use that $\mathbb{E}T'_\emptyset \geq 1/(2\pi'_0)$, since π'_0 is very close to 0 in cases which are of practical interest, and we let μ_π denote the mean of π.

Consider first the Poisson case $\pi_i \propto \beta^i/i!$. By (2.30), $\mathbb{E}T'_\emptyset \geq e^\beta$ is at least exponentially growing in β. For example, if $\beta = 100$ (i.e. $\mu_\pi = 100$),

$$\mathbb{E}T'_\emptyset \geq e^{100} \approx 2.7 \times 10^{43}. \tag{2.31}$$

Consider next the case where $\pi_i \propto q^i\Gamma(i + \alpha)/i!$ is negative binomial as in Example 3, where $\alpha > 0$ and $0 < q < 1$. Let $d_i = i$ and $b_i = (i + \alpha)q$ (as noticed in Section 2.3.5, Q is then non-explosive). By (2.30),

$$\mathbb{E}T'_\emptyset \geq (1 - q)^{-(1+\alpha)}. \tag{2.32}$$

If $(1-q)^{1-\alpha} \leq \alpha q(1+q)$, a better bound is obtained as follows. The probability for a transition $i \to i+1$ in M is $q(i+\alpha)/(q(i+\alpha)+i)$. As i increases, $q(i+\alpha)/(q(i+\alpha)+i)$ decreases towards $p \equiv q/(1+q) < 1/2$. Hence it suffices to consider a random walk $\{S_n : n \in \mathbb{N}_0\}$ starting in $S_0 = N_0$ and with $\mathbb{P}(S_{n+1} - S_n = 1) = 1 - \mathbb{P}(S_{n+1} - S_n = -1) = p$, since we can couple the random walk to M so that $T''_\emptyset \geq \mathcal{T}_0$ where $\mathcal{T}_0 = \inf\{n \in \mathbb{N}_0 : S_n = 0\}$. By Wald's identity, $N_0 = (1-2p)\mathbb{E}(\mathcal{T}_0|N_0)$, and so $\mathbb{E}N_0 \leq (1-2p)\mathbb{E}T''_\emptyset$, i.e.

$$\mathbb{E}T'_\emptyset \geq \alpha q(1+q)/(1-q)^2. \tag{2.33}$$

For instance, in the case of a geometric distribution ($\alpha = 1$), this is a better bound than that in (2.32) when $q > (\sqrt{5}-1)/2 \approx 0.618$, and we obtain

$$\mathbb{E}T'_\emptyset \geq 20100 \tag{2.34}$$

if $q = 100/101$ and $\alpha = 1$ (corresponding to $\mu_\pi = 100$).

This indicates that the running time may be much smaller in the geometric case than in the Poisson case (at least when $\mu_\pi = 100$ in both cases). In the geometric case we expect the distribution of $T'_\emptyset$ to be heavy-tailed so that very long running times may occur. However, in the negative binomial case with $\alpha \leq 1$, a smaller coalescence time is obtained using the algorithm in Section 2.5.2.

2.4.2 Upper and lower processes

A faster perfect simulation procedure can be obtained by constructing upper and lower processes $U^s = \{U^s_t : s \leq t \leq 0\}$ and $L^s = \{L^s_t : s \leq t \leq 0\}$ for $s \leq 0$. These are constructed forward in time as follows. Initially set $U^s_s = D_s$ and $L^s_s = \emptyset$. For $s < t \leq 0$, if $d = D_{t-}$, $u = U^s_{t-}$, and $l = L^s_{t-}$, let

$$\begin{aligned}
D_t = d \quad &\Rightarrow \quad U^s_t = u \quad \text{and} \quad L^s_t = l, \\
D_t = d \backslash \eta_t \quad &\Rightarrow \quad U^s_t = u \backslash \eta_t \quad \text{and} \quad L^s_t = l \backslash \eta_t, \qquad (2.35) \\
D_t = d \cup \xi_t \quad &\Rightarrow \quad U^s_t = \begin{cases} u \cup \xi_t & \text{if } R_t \leq \alpha_{\max}(d,u,l,\xi_t) \\ u & \text{otherwise} \end{cases} \\
&\text{and} \quad L^s_t = \begin{cases} l \cup \xi_t & \text{if } R_t \leq \alpha_{\min}(d,u,l,\xi_t) \\ l & \text{otherwise,} \end{cases} \qquad (2.36)
\end{aligned}$$

where

$$\begin{aligned}
\alpha_{\max}(d,u,l,\xi) &= \max\{\bar{b}(x,\xi) b_{n(x)}/b_{n(d)} : l \subseteq x \subseteq u\}, \qquad (2.37) \\
\alpha_{\min}(d,u,l,\xi) &= \min\{\bar{b}(x,\xi) b_{n(x)}/b_{n(d)} : l \subseteq x \subseteq u\}. \qquad (2.38)
\end{aligned}$$

Other choices of $\alpha_{\max}$ and $\alpha_{\min}$ are possible and may be convenient for computational reasons, cf. the discussion at the end of this section.

The construction in (2.35)–(2.38) ensures the following properties; see also Fig. 2.8. The sandwiching property,

$$L^s_t \subseteq X_t \subseteq U^s_t \subseteq D_t, \quad s \leq t \leq 0, \tag{2.39}$$

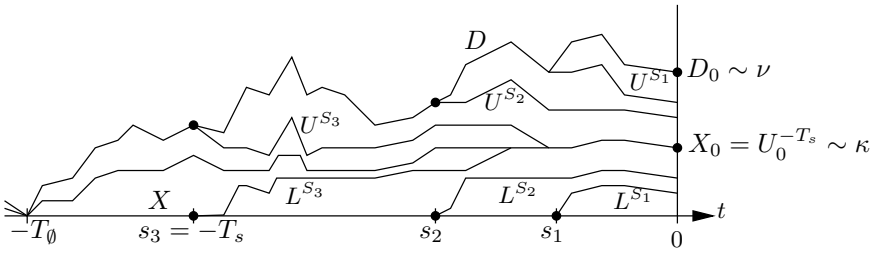

Figure 2.8: Illustration of sandwiching, funnelling, and coalescence properties; see (2.39)–(2.41).

meaning that L^s and U^s are lower and upper bounds on X, and all processes are bounded by D. The funnelling property,

$$L_t^{s'} \subseteq L_t^s \subseteq U_t^s \subseteq U_t^{s'}, \quad s \le s' \le t \le 0, \tag{2.40}$$

meaning that a pair of lower and upper processes started at time $s < 0$ are bounded by any pair of upper and lower processes started at time $s' > s$. And the coalescence property,

$$L_t^s = U_t^s \quad \Rightarrow \quad L_{t'}^s = U_{t'}^s \quad \text{for } s \le t \le t' \le 0, \tag{2.41}$$

meaning that once a pair of upper and lower processes have coalesced, they stay in coalescence, and at time 0 they are equal to $X_0 \sim \kappa$, cf. (2.39).

Now the perfect simulation algorithm works as follows. Pick a sequence of times $\dots s_2 < s_1 < 0$, where $\lim_{i\to\infty} s_i = -\infty$. Then start D in equilibrium at time 0 and generate it backwards until time s_1. For forwards birth times $t \in [s_1, 0]$ in D, generate i.i.d. marks $R_t \sim \text{Uniform}[0,1]$ independent of D. Then generate the upper and lower processes U^{s_1} and L^{s_1} as in (2.35)–(2.36). If U^{s_1} and L^{s_1} are in a common state at time 0 then stop. If not then extend D (together with the marks for forwards births) further backwards from $t = s_1$ to $t = s_2$, construct U^{s_2} and L^{s_2}, and check if $U_0^{s_2} = L_0^{s_2}$. This backwards-forwards step is repeated until the upper and lower chains have coalesced by time 0, i.e. at the coalescence time $-T_s = \sup\{s_i : U_0^{s_i} = L_0^{s_i}\}$. By (2.39), we have $U_0^{-T_s} = X_0$ and hence $U_0^{-T_s} \sim \kappa$. This procedure is illustrated in Fig. 2.8.

Notice that we need only to generate the jump chains of D and the upper and lower processes. Hence it is sensible to let $\{s_i\} \subseteq \{J_{-i}\}$, where $\dots J_{-2} \le J_{-1} < 0$ are the jump times of D before time 0. As $-T_s \le -T \le J_{-n(D_0)}$, where $-T = \sup\{s \le 0 : U_0^s = L_0^s\}$ is the "true coalescence time", it is natural to let $s_1 \le J_{-n(D_0)}$. For efficiency reasons a doubling scheme is usually used (Propp & Wilson 1996), i.e.

$$s_1 = J_{-n}, s_2 = J_{-2n}, s_3 = J_{-4n}, s_4 = J_{-8n}, \dots, \tag{2.42}$$

where $n \in \mathbb{N}$ is a user-specified parameter. Then we write T_n for T_s. Note that $T' \leq T'_n$ for the coalescence times in the jump chains, i.e. T' and T'_s are defined by $-T = J_{-T'}$ and $-T_s = J_{-T'_s}$. The doubling scheme can be refined by letting

$$s_1 = J_{-T'_{\min}}, s_2 = J_{-T'_{\min}-n}, s_3 = J_{-T'_{\min}-2n}, s_4 = J_{-T'_{\min}-4n}, \ldots, \tag{2.43}$$

writing $T_{n,\min}$ for T_s, and where $T'_{\min} = \inf\{k \in \mathbb{N}_0 : D_{J_{-k}} \cap D_0 = \emptyset\}$ specifies the number of jumps D has to go through from the first point in D_0 is born until time zero. Often in applications, $T_n \ll T_\emptyset$, $T'_n \ll T'_\emptyset$ and $T'_{n,\min} \leq T'_n$; this is illustrated later in Example 7.

The calculation of $\alpha_{\max}$ and $\alpha_{\min}$ is particular simple in the following cases. A point processes is attractive respectively repulsive if

$$\bar{b}(x,\xi) \leq \bar{b}(y,\xi) \qquad \text{whenever} \qquad x \subseteq y, \xi \notin y, \tag{2.44}$$
$$\bar{b}(x,\xi) \geq \bar{b}(y,\xi) \qquad \text{whenever} \qquad x \subseteq y, \xi \notin y. \tag{2.45}$$

For example, for the point processes considered in Examples 4 and 5, either (2.44) or (2.45) is satisfied. In the attractive case (2.44) we obtain by (2.21) and (2.37)–(2.38),

$$\alpha_{\max}(d,u,l,\xi) = \bar{b}(u,\xi) b_{n(u)}/b_{n(d)},$$
$$\alpha_{\min}(d,u,l,\xi) = \bar{b}(l,\xi) b_{n(l)}/b_{n(d)}.$$

In the repulsive case (2.45), it is computational convenient to redefine $\alpha_{\max}$ and $\alpha_{\min}$ by

$$\alpha_{\max}(d,u,l,\xi) = \bar{b}(l,\xi) b_{n(u)}/b_{n(d)},$$
$$\alpha_{\min}(d,u,l,\xi) = \bar{b}(u,\xi) b_{n(l)}/b_{n(d)}.$$

Then the perfect simulation algorithm still works as (2.39)–(2.41) hold. Note that it is only in the attractive case that U^s and L^s are individual Markov chains. Observe also that in Fig. 2.4, the point pattern for $\gamma = 0$ is not contained in that for $\gamma = 0.5$. This is possible because the Strauss model is non-attractive.

2.4.3 Clan of ancestors

An alternative perfect simulation algorithm is given by Fernández et al. (2000). We assume that ϕ has finite range of interaction, i.e. there exists an $R < \infty$ such that for any $x \in \Omega$ and $\xi \in S \backslash x$, $\bar{b}(x,\xi) = \bar{b}(x \cap \text{ball}(\xi, R), \xi)$. This is fulfilled in Examples 4 and 5.

The algorithm is based on two central concepts defined as follows. Consider again the coupling construction in Section 2.3.5. When we have a birth $D_t = D_{t-} \cup \xi_t$, the *parents* of $\xi = \xi_t$ is the set $\text{pa}(\xi) = D_{t-} \cap \text{ball}(\xi, R)$ (we suppress in the notation that $\text{pa}(\xi)$ depends on D_{t-}). Whether $X_t = X_{t-} \cup \xi$ or $X_t = X_{t-}$ in (2.24) depends on D_{t-} only through $\text{pa}(\xi)$, or in fact only through those points in $\text{pa}(\xi)$ which have not been accepted earlier in the thinning of D. In

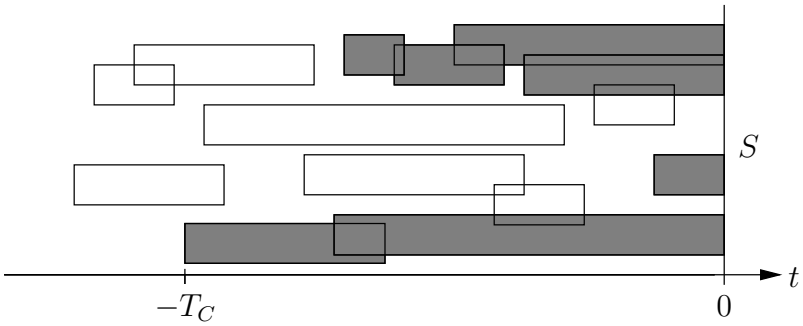

Figure 2.9: Example of a clan of ancestors where shaded blocks represents members of the clan.

this sense whether ξ should be added to X depends not only on its parents, but also on the parents of its parents, and so on. Therefore we define recursively the ith generation ancestors of ξ by $\mathrm{pa}^i(\xi) = \cup_{\eta\in\mathrm{pa}^{i-1}(\xi)}\mathrm{pa}(\eta)$, $i = 2, 3, \ldots$, where $\mathrm{pa}^1(\xi) = \mathrm{pa}(\xi)$ (again we suppress the dependence of D when defining $\mathrm{pa}^i(\xi)$). Furthermore, let $\mathrm{an}(\xi) = \cup_{i\in\mathbb{N}}\mathrm{pa}^i(\xi)$ be the union of all generations of ancestors of ξ. If $\xi \in D_0$, then the points in $\mathrm{an}(\xi)$ are the only points which can have any influence on whether ξ is in X_0 or not. So $C(D_0) = \cup_{\xi\in D_0}\mathrm{an}(\xi) \cup D_0$ is the set of *all* points in D which can have any influence on the configuration of X_0. In other words, if $[-T_C, 0]$ is the time interval in which the points in $C(D_0)$ are living, then $T_C \leq T_\emptyset$, and X_0 is unaffected if we set $X_{-T_C} = \emptyset$ and generate X_t forwards in time $t \geq -T_C$ as usual. We refer to $C(D_0)$ as the *clan of ancestors*. An example of a clan is shown in Fig. 2.9.

Now, briefly the perfect simulation algorithm works as follows (noticing that we only need to generate the relevant jump chains). Let $D_0 \sim \nu$. As noticed above we only need to generate the dominating process D back until time $-T_C$, which may be determined as follows. Initially all points in D_0 are set to be in the clan. Then whenever a member of the clan dies in D (when D is considered backwards in time), e.g. $D_t = D_{t-} \cup \{\xi_t\}$, all points in D_{t-} within distance R of ξ_t are the parents of ξ_t and are therefore added to the clan. When all members of the clan have died, we have found the time T_C and the generation of D stops. Finally, we set $X_{-T_C} = \emptyset$, construct X_t, $t \geq -T_C$, as in Section 2.4.1, and return $X_0 \sim \kappa$.

As noticed, $T_C \leq T_\emptyset$, and it is not hard to see that $T \leq T_C$. Similarly, $T' \leq T'_C \leq T'_\emptyset$, where $T_C = -J_{-T'_C}$. Note that the running time of the algorithm depends only on $\bar{b}$ through R, and no monotonicity properties such as (2.44) and (2.45) are required.

Example 7: Consider a Gibbs hard core process as given by (2.19) with $\gamma = 0$ and hard core parameter $R \geq 0$, letting S be a unit square, and ν a homogeneous Poisson point process of rate $\beta = 100$. Such hard core processes are studied in the contributions by Schmid & Phuong and by Döge in this collection; see

also (Döge 2000, Löwen 2000, Mase, Møller, Stoyan, Waagepetersen & Döge 2001). Fig. 2.10 shows perfect simulations of a hard core process for different values of R.

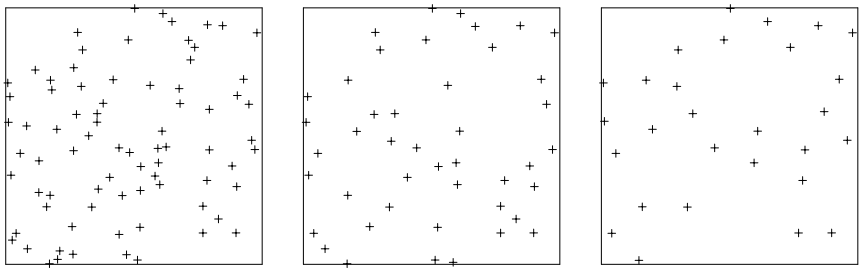

Figure 2.10: Simulation of a Gibbs hard core process (see Example 7) when $R = 0.03, 0.07, 0.11$ (from left to right).

Fig. 2.11 shows the mean coalescence times of T'_C, T'_n, and $T'_{n,\min}$ versus R, when $n = 1$ in the doubling schemes. For all values of R in Fig. 2.11, the mean coalescence times are much smaller than $\mathbb{E}T'_\emptyset \geq e^{100}$, cf. (2.31). For small values of R, the clan method and the refined doubling scheme are almost equal (in the sense that $\mathbb{E}(T'_C) \approx \mathbb{E}(T'_{n,\min})$) and slightly better than the (ordinary) doubling scheme (where the mean coalesce times are about 2/3 smaller). As R increases, T'_n and $T'_{n,\min}$ do not increase as rapidly as T'_C, and the clan method is clearly slower than both doubling schemes for large values of R. This is not so surprising, as an increase in R implies an increase in the number of parents for each point, and hence a larger clan reaching further back in time. The specification of β is of course also crucial: T'_C tends to be smaller than T'_n for small values of β, while the opposite is the case for large values of β.

Further results on the running time for the algorithm based on upper and lower processes are given in (Berthelsen & Møller 2001a).

2.5 Spatial birth-and-catastrophe process

2.5.1 Description and construction

2.5.1.1 The simple case

We start by specifying a simple birth-and-catastrophe process $N = \{N_t : t \geq 0\}$ with generator Q given by $q_{i+1,0} = d_{i+1}$ and $q_{i,i+1} = b_i$, while all other off-diagonal elements are zero. Hence $q_i = b_i + d_i$ for all $i \in \mathbb{N}_0$, setting $d_0 = 0$. Note that d_i now has another interpretation than in Section 2.3.

As in Section 2.2, (i)–(iv) are assumed to hold. For simplicity we assume that $I = \mathbb{N}_0$ and $d_{i+1} > 0$ and $b_i > 0$ for all $i \in \mathbb{N}_0$. Then (ii) is satisfied.

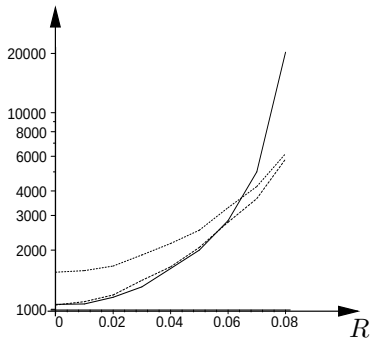

Figure 2.11: Mean coalescence times versus the hard core parameter R for a hard core process. Full line: $\mathbb{E}(T'_C)$; upper dotted line: $\mathbb{E}(T'_n)$; lower dotted line: $\mathbb{E}(T'_{n,\min})$. Each mean value is estimated from 500 independent perfect simulations.

Further (iii) is equivalent to that both

$$\pi_{i+1} = \pi_0 \frac{b_0}{q_1}\frac{b_1}{q_2}\cdots\frac{b_i}{q_{i+1}}, \quad i \in \mathbb{N}_0, \tag{2.46}$$

and

$$\lim_{n\to\infty} \frac{b_0 \cdots b_i}{q_1 \cdots q_i} = 0 \tag{2.47}$$

are satisfied. To see this, rewrite $\pi Q = 0$ as

$$\pi_0 b_0 = \sum_{i=1}^{\infty} \pi_i d_i \quad \text{and} \quad \pi_i b_i = \pi_{i+1} q_{i+1} \quad \text{for all } i \in \mathbb{N}_0. \tag{2.48}$$

From the latter equation, using induction, we obtain (2.46). Combining the equations in (2.48),

$$\pi_0 b_0 = \sum_{i=1}^{\infty} \pi_{i-1} b_{i-1} - \pi_i b_i = \pi_0 b_0 - \lim_{i\to\infty} \pi_i b_i,$$

whereby (2.47) is obtained.

2.5.1.2 The spatial case

We next construct a spatial birth-and-catastrophe process $X = \{X_t : t \geq 0\}$, proceeding along similar lines as in Sections 2.3.3–2.3.5. The process is specified by two functions $\beta : \Omega \times S \to [0, \infty)$ and $\delta : \Omega \to [0, \infty)$, where for any $x \in \Omega$, $\beta(x, \xi)$ is the birth rate at which a point $\xi \in S$ is added to x, and if $x \neq \emptyset$, $\delta(x)$ is the catastrophe rate a which all points are deleted from x. As in Sections 2.2

and 2.3.3, let $J_1 < J_2 < \dots$ be the jump times, $H_1, H_2, \dots$ the holding times, and $Y = \{Y_n : n \in \mathbb{N}_0\}$ the jump chain of X with $Y_0 = X_0$. The conditional distribution of H_n given $H_1, \dots, H_{n-1}, Y_0, \dots, Y_{n-1} = x$ is $\mathrm{Exp}(B(x) + \delta(x))$, where $B(x) = \int_S \beta(x, \xi)\lambda(d\xi)$ is the total birth rate (assuming $B(x) < \infty$) and we set $\delta(\emptyset) = 0$. If we also condition on H_n, the probability for a birth at time J_n is $B(x)/(B(x) + \delta(x))$. If we further condition on that a birth happens at time J_n, then $Y_n = x \cup \xi$ where ξ has density $\beta(x, \cdot)/B(x)$. If instead a catastrophe happens at time J_n, then $Y_n = \emptyset$.

We assume that β and δ are related to the rates for N by

$$\beta(x, \xi) = b_{n(x)}\bar{b}(x, \xi) \quad \text{and} \quad \delta(x) = d_{n(x)}\bar{d}(x), \tag{2.49}$$

where the constraints $0 \le \bar{b} \le 1$ and $\bar{d} \ge 1$ are satisfied. Then, by (2.49), $B(x) \le b_{n(x)}$ and $\delta(x) \ge d_{n(x)}$. Furthermore, we assume that $d_1 \ge d_2 \ge \dots$. Then X and N can be coupled so that N dominates $n(X)$ as in (2.11). Briefly, (X, N) is a jump process where a transition from (x, n) to (x', n') happens with a rate given as follows, where $m = n(x) \le n$:

b_n for $(x', n') = (x, n + 1)$ and $m < n$,

$\beta(x, \xi)$ for $(x', n') = (x \cup \xi, n)$ and $m < n$,

$\beta(x, \xi)$ for $(x', n') = (x \cup \xi, n + 1)$ and $m = n$,

$b_n - B(x)$ for $(x', n') = (x, n + 1)$ and $m = n$,

d_n for $(x', n') = (\emptyset, 0)$ and $m \le n$,

$\delta(x) - d_n$ for $(x', n') = (\emptyset, n)$ and $m \le n$.

As in Section 2.3.3 we obtain the following: X is non-explosive; $\emptyset$ is an ergodic atom at which X regenerates; as in (2.12), X_t converges in distribution towards an equilibrium distribution κ, say, independent of the initial state of X_0 as $t \to \infty$; and κ has a support given by (2.14).

The equilibrium distribution is uniquely characterised by the equations

$$B(\emptyset)\kappa(\emptyset) = \int_{x \ne \emptyset} \delta(x)\kappa(dx) \tag{2.50}$$

and

$$\int_{F_i} [B(x) + \delta(x)]\kappa(dx) = \iint_{x \cup \xi \in F_i} \beta(x, \xi)\lambda(d\xi)\kappa(dx) \tag{2.51}$$

for all $i \in \mathbb{N}$ and events $F_i \subseteq \Omega_i$ (this follows from (2.56) in Section 2.6). If κ has an unnormalised density ϕ with respect to ν given by (2.15), then (2.50) and (2.51) become

$$b_0\pi_0\phi(\emptyset) \int \bar{b}(\emptyset, \xi)\lambda(d\xi) = \sum_{i=1}^{\infty} d_i\pi_i \int \bar{d}(x)\phi(x)\lambda_i(dx) \tag{2.52}$$

and

$$\int_{F_i} \pi_i \phi(x) \left(b_i \int \bar{b}(x,\xi)\lambda(d\xi) + d_i \bar{d}(x) \right) \lambda_i(dx) = \iint_{x \cup \xi \in F_i} \pi_{i-1} b_{i-1} \bar{b}(x,\xi) \phi(x) \lambda_{i-1}(dx). \quad (2.53)$$

Using (2.48), then (2.52) and (2.53) are easily seen to be satisfied if, for example, ϕ is non-increasing as in (2.17), $\bar{b}$ is the Papangelou intensity as in (2.18), and

$$\bar{d}(x) = 1 + \left(b_{n(x)} - B(x) \right) / d_{n(x)} \quad \text{for } x \neq \emptyset.$$

In the sequel we assume that $\bar{b}$ and $\bar{d}$ are defined in this way for a given unnormalised density ϕ which is non-increasing. These assumptions imply that the total rates in N and X agree in the sense that for any $x \in \Omega$, $B(x) + \delta(x) = q_{n(x)}$. In other words, if $X_t = x$, the waiting time τ to the next jump is $\mathrm{Exp}(q_{n(x)})$-distributed, and we may generate $\xi \sim \lambda$ and return $X_{t+\tau} = x \cup \xi$ with probability $\beta(x,\xi)/q_{n(x)}$, and return $X_{t+\tau} = \emptyset$ otherwise.

2.5.2 Perfect simulation

In this section we show that the simple perfect simulation procedure in Section 2.4.1 is feasible when the dominating simple jump process is of the irreversible type presented in Section 2.5.1. In addition to the assumptions in Section 2.5.1, we assume for simplicity that $q_i \equiv 1$ is constant for all $i \in \mathbb{N}$ and $b_0 \leq b_1$. Equivalently, by (2.48), $b_i = \pi_{i+1}/\pi_i \leq 1$ is non-decreasing for $i \in \mathbb{N}_0$, and $d_i = 1 - b_i$ for $i \in \mathbb{N}$. Hence the Poisson case $\pi_i \propto \beta^i / i!$ is excluded, since $\pi_{i+1}/\pi_i = \beta/(i+1)$ is decreasing in i. If $\pi_i \propto \Gamma(i+\alpha) q^i / i!$ is negative binomial (Example 3), $\pi_{i+1}/\pi_i = q(i+\alpha)/(i+1) \leq 1$ is non-decreasing if and only if $\alpha \leq 1$.

Now, a coupling construction is easily specified: Suppose N_t is generated for all times $t \geq 0$, and for all jump times t of N_t, there are generated mutually independent points $\xi_t \sim \lambda$ and numbers $R_t \sim \mathrm{Uniform}[0,1]$ (independent of N). Let $n(X_0) \leq N_0$ and generate X_t forwards in time $t > 0$ as follows. Suppose that $X_{t-} = x$ and $N_{t-} = j$ with $n(x) = i \leq j$. If $N_t = j$ is unchanged, then $X_t = x$ is unchanged, while

$$N_t \neq j \Rightarrow X_t = \begin{cases} x \cup \xi_t & \text{if } N_t = j+1 \text{ and } R_t \leq \bar{b}(x,\xi_t) b_i / b_j \\ \emptyset & \text{otherwise.} \end{cases} \quad (2.54)$$

Since $\{b_j\}$ is non-decreasing, we obtain a spatial birth-and-catastrophe process with rates as required, and by induction, $N_t \geq n(X_t)$ for all $t \geq 0$.

For perfect simulation, we let $N_0 \sim \pi$ and imagine that N_t is generated backwards in time $t \leq 0$ until it reaches zero, i.e. at time $-\tilde{T}_\emptyset \equiv \sup\{t \leq 0 : N_t = 0\}$. The number of jumps in $[-\tilde{T}_\emptyset, 0]$ is simply given by N_0, since N_t can

only move one step up when considered forwards in time between $-\tilde{T}_\emptyset$ and 0. Hence, in accordance with (2.54), we let $Y_0 = \emptyset$ and generate Y_i, $i = 1, \ldots, N_0$, by

$$Y_i = \begin{cases} Y_{i-1} \cup \xi_i & \text{if } R_i \leq \bar{b}(Y_{i-1}, \xi_i) b_{n(Y_{i-1})}/b_{i-1} \\ \emptyset & \text{otherwise,} \end{cases} \tag{2.55}$$

where $\xi_i \sim \lambda$ and $R_i \sim \text{Uniform}[0,1]$. Since $Y_0, \ldots, Y_{N_0}$ can be considered as the part of the jump chain of X_t ("started in the infinite past") where $t \in [-\tilde{T}_\emptyset, 0]$, we conclude that $Y_{N_0} \sim \kappa$ (a formal argument follows exactly the same lines as in Section 2.4.1).

So perfect simulation is very simple:

(I) generate $N_0 \sim \pi$;

(II) generate the jump chain $Y_0, \ldots, Y_{N_0}$;

(III) return $Y_{N_0} \sim \kappa$.

The number of jumps is given by $N_0 \sim \pi$. This should be compared to the case in Example 6: If $\pi_i \propto q^i$ is geometric, its mean $q/(1-q)$ is $(1+q)/(1-q)$ times smaller than the lower bound on $\mathbb{E}T'_\emptyset$ in (2.33). Hence it is much faster; for example, $(1+q)/(1-q) = 201$ if $\mu_\pi = 100$, cf. (2.34).

Unless $\bar{b}$ is very close to 1 (corresponding to a weak interaction), we might by (2.55) expect $n(Y_{N_0})$ to be small. This is in accordance with the observations in Example 4, and it explains why the perfect simulation algorithm is so fast. Furthermore, in contrast to the algorithms in Section 2.4, there is no need for storing any information: our perfect sampler is a simple example of a so called read-once algorithm (Wilson 2000a).

2.6 The general case

In a general setting, a spatial jump process $X = \{X_t : t \geq 0\}$ may be specified by functions $\beta, \delta : \Omega \times \Omega \to [0, \infty)$, where for any $x \in \Omega$, $\beta(x, y)$ is the rate at which a point configuration $y \in \Omega$ is added to x, and if $z = x \cup y \in \Omega$, $\delta(x, y)$ is the rate at which y is deleted from z; in order to ensure that the process really jumps, we set $\beta(x, \emptyset) = 0$ and $\delta(x, \emptyset) = 0$. Let further $B(x) = \int \beta(x, y)\nu(dy)$ and $D(z) = \sum_{x \subseteq z} \delta(x, z \backslash x)$, where ν is defined as in (2.15) with respect to a given probability density function $\pi = \{\pi_i\}$ on $\mathbb{N}_0$. Then X can be defined in terms of its jump times, holding times, and jump chain in the same way as in Section 2.3.3. Especially, for the jump chain $\{Y_n\}$, in case of a transition $Y_n = x \cup y$ given that $Y_{n-1} = x$, y has density $\beta(x, \cdot)/B(x)$ with respect to ν (assuming $B(x) < \infty$), and in case of a transition $Y_n = x$ given that $Y_{n-1} = z = x \cup y$, y has been selected with probability $\delta(x, y)/D(z)$.

Let us, as before, restrict attention to the non-explosive case. For instance, X is non-explosive if $B(x) + D(x)$ is uniformly bounded. By Proposition 8.1

in Preston (1977), a probability measure κ on Ω is invariant if and only if

$$\int_{F_i} [B(x) + D(x)]\, \kappa(dx) = \iint_{x \cup y \in F_i} \beta(x, y)\nu(dy)\kappa(dx) + \int \sum_{x \subseteq z : x \in F_i} \delta(x, z \backslash x)\kappa(dz) \tag{2.56}$$

for all $i \in \mathbb{N}_0$ and events $F_i \subseteq \Omega_i$. Here invariance means that $X_t \sim \kappa$ for all $t \geq 0$ if $X_0 \sim \kappa$. In particular, if X converges in distribution towards κ, then κ is the unique invariant distribution. For example, assume that Q and π are in balance, cf. (2.1), and

$$\beta(x, y) = q_{n(x),n(x)+n(y)}\bar{b}(x, y)$$

and

$$\delta(x, y) = q_{n(x)+n(y),n(x)}\bar{d}(x, y) / \binom{n(x) + n(y)}{n(x)}.$$

Then, if ϕ is an unnormalised density for κ with respect to ν, there is a detailed balance condition:

$$\phi(x)\bar{b}(x, y) = \phi(x \cup y)\bar{d}(x, y) \quad \text{for all } x, y \in \Omega. \tag{2.57}$$

It is straightforwardly verified that this implies both (2.56) and reversibility of X.

In Sections 2.3 and 2.5.1 we demonstrated that a coupling of X to a simple jump process N is very useful. However, at the present level of generality it seems difficult to specify a coupling so that N dominates $n(X)$. Probably we need to consider this problem case by case; indeed the coupling constructions in Sections 2.3 and 2.5.1 depend much on the choice of β and δ. For making perfect simulations from a given invariant distribution κ, we may simply try to use one of the algorithms in Sections 2.4 and 2.5.2.

Example 8: Let $\pi_i \propto \beta^i/i!$ be Poisson, where $\beta > 0$,

$$q_{i,i+j} = p^i(\beta(1 - p))^j/j!, \quad q_{i+j,i} = \binom{i+j}{i} p^i(1 - p)^j,$$

where $0 < p < 1$, and

$$\bar{d} \equiv 1, \quad \bar{b}(x, y) = \phi(x \cup y)/\phi(y),$$

with $\phi > 0$ non-increasing. Then the detailed balance conditions (2.1) and (2.57) are satisfied, and X develops as follows. Suppose that $X_t = x$. Then

(I) with rate $p^{n(x)}e^{\beta(1-p)}$, first propose a birth $x \to x \cup y$, drawn from a Poisson process with intensity measure $\beta(1 - p)\lambda$, next with probability $\phi(x \cup y)/\phi(x)$ accept $x \cup y$, and retain x otherwise;

(II) with rate 1 make an independent thinning of the points in x, where each point is retained with probability p.

Note that nothing may happen with rate $2p^{n(x)}$ (namely if $y = \emptyset$ in (I) or all points are retained in (II)). Observe also that y is independent of x in (I), and the retention probability p does not depend on x in (II).

Both N and X are non-explosive as $q_i = p^i e^{\beta(1-p)} + 1 - 2p^i$ and $B(x) + D(x) \leq q_{n(x)}$ are bounded. It is easily seen that $q_{i,i+j} \geq q_{i-k,i+j}$ for all integers $0 < k \leq i$ and $j \geq 1$ if and only if $p \geq \beta/(2+\beta)$. Hence, if $p \geq \beta/(2+\beta)$, we obtain that N dominates X by specifying a jump process (X, N) as follows. The rate for a jump $(x, n) \to (x', n')$ is as follows:

$$\beta(x, y) \quad \text{for } (x', n') = (x \cup y, n(x \cup y)) \text{ and } n < n',$$

$$q_{n,n'} - \textstyle\int_{n(x \cup y) = n'} \beta(x, y)\nu(dy) \quad \text{for } x' = x \text{ and } n < n',$$

$$\beta(x, y) \quad \text{for } (x', n') = (x \cup y, n) \text{ and } n(x \cup y) \leq n,$$

$$q_{n,n'} \quad \text{for } x' = x \backslash y \text{ and } n' \leq n - n(y).$$

Because of this domination, we have the usual ergodicity properties: $\emptyset$ is an ergodic atom at which X regenerates, and X converges in distribution towards κ (independent of the initial state of X_0).

It is tempting to try to extend this coupling construction to a dominating spatial jump process $D \supseteq X$, but this idea fails because of the accept-rejection step (I). However, we can still simulate X_0 from its equilibrium distribution as any of the perfect simulation algorithms in Section 2.4 apply (for any $\beta > 0$ and $0 < p < 1$).

Acknowledgement

This research will be a part of KKB's PhD dissertation. JM was supported by the European Union's research network "Statistical and Computational Methods for the Analysis of Spatial Data, ERB-FMRX-CT96-0096", by the Centre for Mathematical Physics and Stochastics (MaPhySto), funded by a grant from the Danish National Research Foundation, and by the Danish Natural Science Research Council. We thank Bjarne Højgaard for helpful discussions.

Appendix: notation index

Symbol	Explanation	Page
S	The state space for "points"	31, 37
X (X_t)	In general: spatial jump process (at time t) In Sections 2.3–2.4: spatial birth-and-death process In Section 2.5: spatial birth-and-catastrophe process	34
N (N_t)	Simple jump process (at time t)	34
Q (q_{ij},q_i)	Generator of N (with elements q_{ij}, where $q_i = -q_{ii}$)	34
π (π_i)	The invariant probability density function for N (with probabilities π_i for $i \in I$)	34
I	The support for π	34
b_i, d_i	In Section 2.3.1: birth and death rates for N In Section 2.5.1.1: birth and catastrophe rates for N	36, 51
β	If π is a Poisson distribution, β is the mean for π and $-\ln\beta$ is the chemical activity for $Z \sim \phi$	36, 40
q, α	Parameters for the geometric/negative binomial distribution defined in Example 2 and 3	36
$\Omega = \cup_{i=0}^{\infty}\Omega_i$	The state space of X_t	37
Ω_i	The set of finite point configurations confined in S and of cardinality i	37
$n(x)$	The number of points in a finite point configuration $x \in \Omega$	37
$\emptyset$	The empty point configuration	37
λ	Probability measure on S	37
$\beta(\cdot,\cdot)$, $\delta(\cdot,\cdot)$	In Section 2.3.3: birth and death rates for X. In Section 2.5.1.2: birth and catastrophe rates for X In Section 2.6: rates for adding/deleting point configurations to/from X	37, 52, 55
$\bar{b}(\cdot,\cdot)$, $\bar{d}(\cdot,\cdot)$	Functions used for relating $\beta(\cdot,\cdot)$, $\delta(\cdot,\cdot)$ to b_i, d_i	38
κ	Invariant distribution for X	39
ϕ	Unnormalised density for κ w.r.t. ν	39

Symbol	Explanation	Page
ν	Probability measure on Ω (used as an reference measure)	39
γ, R	In Example 4: the interaction parameter and interaction radius for a Strauss process In Example 5: the parameters for an area interaction process	40
D, (D_t)	The dominating process (at time t)	42
$T_\emptyset$	$-T_\emptyset$ is the first time before time 0 that D_t enters $\emptyset$	45
$T'_\emptyset$	The number of jumps D makes on $[-T_\emptyset, 0]$	45
μ_π	The mean for the distribution π	46
U^s (U^s_t), L^s (L^s_t)	Upper/lower processes started at time $s \leq 0$ (and considered at time t)	47
T (T')	True coalescence time for upper and lower processes (or for their corresponding jump chains)	48
T_n (T'_n)	Coalescence time when a doubling scheme is used for upper and lower processes (or for their corresponding jump chains)	49
$T_{n,\min}$ $(T'_{n,\min})$	Coalescence time when a refined doubling scheme is used for upper and lower processes (or for their corresponding jump chains)	49

Bibliography

Asmussen, S. (1987). *Applied Probability and Queues*, Wiley, Chichester.

Baddeley, A. & van Lieshout, M. N. M. (1995). Area-interaction Point Processes, *Ann. Inst. Statist. Math.* **46**: 601–619.

Berthelsen, K. K. & Møller, J. (2002). A primer on perfect simulation for spatial point processes, *Bull. Braz. Math. Soc.* **33**: 351–367.

Berthelsen, K. K. & Møller, J. (2003). Likelihood and non-parametric Bayesian MCMC inference for spatial point processes based on perfect simulation and path sampling. *Scand. J. Statist.* To appear.

Daley, D. J. & Vere-Jones, D. (1988). *An Introduction to Point Processes*, Springer, Berlin.

Döge, G. (2000). Grand canonical simulations of hard-disk systems, *in* K. R. Mecke & D. Stoyan (eds), *Statistical Physics and Spatial Statistics*, Vol. 554 of *Lecture Notes in Physics*, Springer, Berlin, pp. 373–393.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, Vol. 2, 2nd edn, John Wiley and Sons Inc., New York.

Fernández, R., Ferrari, P. A. & Garcia, N. L. (2000). Perfect simulation for interacting point processes, loss networks and Ising models. Submitted to *Stoch. Process. Appl.*

Georgii, H.-O. (2000). Phase transitions and percolation in Gibbsian particle models, *in* K. R. Mecke & D. Stoyan (eds), *Statistical Physics and Spatial Statistics*, Vol. 554 of *Lecture Notes in Physics*, Springer, Berlin, pp. 267–294.

Geyer, C. J. (1999). Likelihood inference for spatial processes, *in* O. E. Barndorff-Nielsen, W. Kendall & M. van Lieshout (eds), *Stochastic Geometry, Likelihood and Computation*, Chapman & Hall, pp. 79–140.

Häggström, O., van Lieshout, M. N. M. & Møller, J. (1999). Characterization results and Markov chain Monte Carlo algorithms including exact simulation for some spatial point processes, *Bernoulli* **5**: 641–658.

Kallenberg, O. (1984). An informal guide to the theory of conditioning in point processes, *Int. Statist. Rev.* **52**: 151–164.

Kelly, F. P. & Ripley, B. D. (1976). A note on Strauss's model for clustering, *Biometrika* **63**: 357–360.

Kendall, W. S. (1998). Perfect simulation for the area-interaction point process, *in* L. Accardi & C. Heyde (eds), *Probability Towards 2000*, Springer, New York, pp. 218–234.

Kendall, W. S. & Møller, J. (2000). Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes, *Adv. Appl. Prob.* **32**: 844–865.

Lieshout, M. N. M. van (2000). *An Markov Point Processes and their Application*, Imperial College Press, London.

Löwen, H. (2000). Fun with hard spheres, *in* K. R. Mecke & D. Stoyan (eds), *Statistical Physics and Spatial Statistics*, Vol. 554 of *Lecture Notes in Physics*, Springer, Berlin, pp. 295–331.

Mase, S., Møller, J., Stoyan, D., Waagepetersen, R. & Döge, G. (2001). Packing densities and simulated tempering for hard core Gibbs processes, *Ann. Inst. Statist. Math.* **53**: 661–680.

Mecke, K. (2000). Additivity, convexity and beyond: applications of Minkowski functionals in statistical physics, *in* K. R. Mecke & D. Stoyan (eds), *Statistical Physics and Spatial Statistics*, Vol. 554 of *Lecture Notes in Physics*, Springer, Berlin, pp. 111–184.

Meyn, S. P. & Tweedie, R. L. (1994). *Markov Chains and Stochastic Stability*, Springer-Verlag, London.

Møller, J. (1989). On the rate of convergence of spatial birth-and-death processes, *Ann. Inst. Statist. Math.* **41**: 565–581.

Møller, J. (2001). A review of perfect simulation in stochastic geometry, *in* I. V. Basawa, C. C. Heyde & R. L. Taylor (eds), *Selected Proceedings of the Symposium on Inference for Stochastic Processes*, IMS Lecture Notes & Monographs Series. To appear.

Møller, J. & Schladitz, K. (1999). Extensions of Fill's algorithm for perfect simulation, *J. Roy. Statist. Soc. Ser B* **61**: 1–15.

Norris, J. R. (1997). *Markov Chains*, Cambridge University Press, New York.

Preston, C. (1977). Spatial birth-and-death processes, *Bull. Int. Statist. Inst.* **46**: 371–391.

Propp, J. G. & Wilson, D. B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics, *Random Structures and Algorithms* **9**: 223–252.

Reuter, G. E. H. & Ledermann, W. (1953). On the differential equations for the transition probabilities of Markov processes with enumerably many states, *Proc. Camb. Phil. Soc.* **49**: 247–262.

Ripley, B. D. (1977). Modelling spatial patterns (with discussion), *J. Roy. Statist. Soc. Ser. B* **39**: 172–212.

Ruelle, D. (1969). *Statistical Mechanis: Rigorous Results*, W. A. Benjamin, Reading, Massachusetts.

Strauss, D. J. (1975). A model for clustering, *Biometrika* **62**: 467–475.

Thönnes, E. (1999). Perfect simulation of some point processes for the impatient user, *Adv. Appl. Prob.* **31**: 69–87.

Thönnes, E. (2000). A primer on perfect simulation, *in* K. R. Mecke & D. Stoyan (eds), *Statistical Physics and Spatial Statistics*, Vol. 554 of *Lecture Notes in Physics*, Springer, Berlin, pp. 349–378.

Widom, B. & Rowlinson, J. S. (1970). A new model for the study of liquid-vapor phase transitions, *J. Chem. Phys.* **52**: 1670–1684.

Wilson, D. B. (2000). How to couple from the past using a read-once source of randomness, *Random Structures and Algorithms* **16**: 85–113.

3 A primer on perfect simulation for spatial point processes

Kasper K. Berthelsen and Jesper Møller

Department of Mathematical Sciences, Aalborg University,
Fredrik Bajers Vej 7G, DK-9220 Aalborg, Denmark

Abstract. This primer provides a self-contained exposition of the case where spatial birth-and-death processes are used for perfect simulation of locally stable point processes. Particularly, a simple dominating coupling from the past (CFTP) algorithm and the CFTP algorithms introduced in Kendall (1998), Kendall & Møller (2000), and Fernández et al. (2000) are studied. Some empirical results for the algorithms are discussed.

Keywords: Coupling from the past (CFTP), clans of ancestors, dominated CFTP, exact simulation, local stability, spatial birth-and-death process, Strauss process.

Mathematical subject classification: Primary 60G55, 62M30, 65C05; Secondary: 60D05, 60K35, 68U20.

3.1 Introduction

One of the most exciting and important recent developments in Markov chain Monte Carlo (MCMC) is perfect or exact simulation. Following the seminal work by Propp & Wilson (1996) many new perfect simulation ideas have appeared, particularly for spatial point processes, cf. the survey in Møller (2001); see also Wilson's web site (http://dimacs.rutgers.edu/∼dwilson/exact.html). The aims of this paper are to review and compare the performance of some perfect simulation algorithms which apply on a rather general class of point processes, viz. locally stable point processes. For simplicity, apart from Section 3.8, we consider only finite point processes.

We focus on algorithms based on dominated (or horizontal) coupling from the past (CFTP) using spatial birth-and-death processes; alternative and efficient perfect samplers have been developed for some special models, see Møller (2001) and the references therein. In Kendall & Møller (2000) dominated CFTP is treated in a general context and applied on locally stable point processes using either spatial birth-and-death processes or a Metropolis-Hastings algorithm. In this paper we give an alternative and self-contained exposition of the case where

spatial birth-and-death processes are used. A spatial birth-and-death process is a continuous time Markov process where each transition consists in either adding a new point to the process (a birth) or deleting an existing point from the process (a death). Background material on spatial birth-and-death processes can be found in Preston (1977) and Møller (1989), but it is not needed in the present paper. Extensions of the algorithms considered in this paper are given in Berthelsen & Møller (2001b) using spatial jump processes. Another extension which is not treated in this paper, is Wilson's (2000a) read-once version of CFTP. This algorithm applies also on locally stable point processes, and it drastically reduces the storage requirements.

The paper is organized as follows. Section 3.2 describes the setting for spatial point processes used in this paper, and it is explained what is meant by local stability. Section 3.3 specifies a coupling construction which is underlying the perfect samplers considered later. Section 3.4 discusses a very simple perfect simulation algorithm, and we show that it is too slow for practical purposes. Section 3.5 describes a more efficient algorithm based on so-called upper and lower processes (Kendall 1998, Kendall & Møller 2000). Section 3.6 describes an alternative algorithm using so-called clans of ancestors (Fernández et al. 2000). Section 3.7 discusses some empirical findings for the various perfect simulation algorithms. Section 3.8 concludes with some comments on extensions to infinite point processes.

3.2 Background

Throughout this paper we consider a fairly general setting for a spatial point process χ defined on a space S, equipped with a σ-algebra $\mathcal{B}$ which contains all singleton sets, and a diffuse probability measure λ, i.e. $\{\xi\} \in \mathcal{B}$ and $\lambda(\{\xi\}) = 0$ for all $\xi \in S$. For simplicity we assume χ to be a finite subset of S, though everything in the sequel easily extend to the case where χ is allowed to have multiple points and λ is not necessarily diffuse.

The state space of χ is the set of all finite point configurations $\Omega = \bigcup_{i=0}^{\infty}\{x \subseteq S : n(x) = i\}$, where $n(x)$ denotes the number of points in x; for $i = 0$ we have the empty point configuration $x = \emptyset$. We equip Ω with the smallest σ-algebra making the mappings $n_B(x) = n(x \cap B)$ measurable for all $B \in \mathcal{B}$. Further, ν denotes a Poisson point process on S with intensity measure $\beta\lambda$, where $\beta > 0$ is a parameter. In other words, if χ follows ν, then $n(\chi)$ is Poisson distributed with mean β, and conditionally on $n(\chi) = i$, the i points in χ are independent and each point has distribution λ. Specifically one may think of $S = [0,1]^2$ as the unit square, $\mathcal{B}$ as the Borel sets, and λ as the uniform distribution, in which case ν is a standard Poisson process. However, our general setting covers many other cases, including situations where χ can be interpret as a multitype or marked point process, see e.g. Baddeley & Møller (1989) and van Lieshout (2000) .

We assume that the distribution of χ is specified by an unnormalized density

ϕ with respect to ν, so that ϕ is non-increasing in the following sense:

$$\phi(x \cup \xi) \leq \phi(x) \quad \text{for all } x \in S \text{ and } \xi \in S \setminus x \tag{3.1}$$

(we abuse the notation and write $x \cup \xi$ for $x \cup \{\xi\}$, $x \setminus \eta$ for $x \setminus \{\eta\}$, etc., when $x \in \Omega$, $\xi \in S \setminus x$, $\eta \in x$). This condition implies integrability of ϕ with respect to ν. Particularly, (3.1) is needed for the perfect simulation algorithms considered in this paper.

For a moment consider any unnormalized density ϕ with respect to ν. Local stability of ϕ means that for some constant $K > 0$ and all $x \in \Omega$ and all $\xi \in S \setminus x$,

$$\phi(x \cup \xi) \leq K\phi(x) \tag{3.2}$$

(Ruelle 1969). This is a basic assumption in many papers: for example, Geyer (1999) establishes geometric ergodicity of a birth-death type Metropolis-Hastings algorithm for locally stable point processes (Geyer & Møller 1994); and Kendall & Møller (2000) show that it is a sufficient condition for applying dominated CFTP based on spatial birth-and-death processes and Metropolis-Hastings algorithms. Local stability is in fact a rather weak condition satisfied by most models considered in the statistical literature on spatial point processes, cf. the discussion in Kendall & Møller (2000). The concept of local stability is extended in Berthelsen & Møller (2001b) to cases where the dominating measure ν is not necessary a Poisson process.

As K can be absorbed into the parameter β we may without loss of generality set $K = 1$ in (3.2), whereby (3.1) is obtained. Below we consider just two examples where (3.1) is satisfied.

Example 1: Suppose that λ is the uniform distribution on $S = [0,1]^2$ and

$$\phi(x) = \gamma^{s_R(x)} \tag{3.3}$$

taking $0^0 = 1$, where $s_R(x) = \sum_{\{\xi,\eta\} \subseteq x} \mathbf{1}[||\xi - \eta|| \leq R]$ is the number of R-close pairs of points in x, and where $0 \leq \gamma \leq 1$ and $R > 0$ are parameters. This specifies a Strauss process on the unit square (Strauss 1975, Kelly & Ripley 1976). Clearly, ϕ is locally stable.

Example 2: Let S and λ be specified as in Example 1, but let now

$$\phi(x) = \gamma^{-\lambda(U_x)}$$

where $U_x = \cup_{\xi \in x} \text{ball}(\xi, R)$ is the union of closed balls with centers $\xi \in x$ and of radius R, where $R > 0$ and $\gamma > 0$ are parameters. This is an area interaction point process (Widom & Rowlinson 1970, Baddeley & van Lieshout 1995). The process is said to be attractive for $\gamma > 1$, and repulsive for $\gamma < 1$, since

$$\phi(x \cup \xi)/\phi(x) = \gamma^{-\lambda(U_{x \cup \xi} \setminus U_x)} \tag{3.4}$$

is increasing ($\gamma > 1$) or decreasing ($\gamma < 1$) in x. It follows from (3.4) that (3.1) holds in the attractive case, but not in the repulsive case. If $\gamma < 1$ we therefore

redefine ν as a Poisson process with intensity measure $(\beta/\gamma^{\pi R^2})\lambda$, and redefine ϕ by

$$\phi(x) = \gamma^{n(x)\pi R^2 - \lambda(U_x)}.$$

Then (3.1) is satisfied.

3.3 Coupling construction

Below we construct two time-stationary and reversible spatial birth-and-death processes $X = \{X_t : t \in \mathbb{R}\}$ and $D = \{D_t : t \in \mathbb{R}\}$ with equilibrium distributions given by $X_t \sim \phi$ (with respect to ν) and $D_t \sim \nu$. The two processes are coupled so that D dominates X in the sense that

$$X_t \subseteq D_t \quad \text{for all } t \in \mathbb{R}. \tag{3.5}$$

This is obtained by letting (D, X) be a continuous time Markov processes with the following types of transitions: either a new point is added to both D and X, or a birth happens in D but not in X, or a point in X is deleted from both D and X, or a point in D but not in X is deleted. The coupling construction is underlying the perfect samplers in Sections 3.4– 3.6.

We first specify how D_t can be generated forwards in time $t \geq 0$. For any $x \in \Omega$ and $t \geq 0$, if we condition on that $D_t = x$, and τ is the waiting time for the next transition in D after time t, then

- τ is exponentially distributed with mean $1/(\beta + n(x))$;
- with probability $\beta/(\beta + n(x))$ a birth happens in D at time $t + \tau$: draw a point $\xi_{t+\tau} \sim \lambda$ and set $D_{t+\tau} = x \cup \xi_{t+\tau}$ — for later use in the coupling construction, generate also a "mark" $R_{t+\tau} \sim \text{Uniform}[0, 1]$;
- else a death happens in D at time $t + \tau$: draw randomly uniformly a point $\eta_{t+\tau}$ from x and set $D_{t+\tau} = x \setminus \eta_{t+\tau}$.

Furthermore, the conditional distributions of τ, the event of a birth or death, and the generation of either $(\xi, R_{t+\tau})$ or η are assumed to be mutually independent and independent of the previous history given $D_t = x$. In other words, a birth of a new point in D happens with rate β and follows the distribution λ, each point in D dies with rate 1, and births and deaths in D are independent events.

It is easily verified that $\{D_t : t \geq 0\}$ is reversible with invariant distribution ν, and all the marks associated to the birth times are mutually independent and independent of $\{D_t : t \geq 0\}$. Hence we can easily start in equilibrium $D_0 \sim \nu$, and by reversibility, D_t is easily generated backwards in time $t < 0$ together with the associated marks for (forwards) births $D_t = D_{t-} \cup \xi_t$, where $t-$ refers to the situation just before time t. Moreover, it is not hard to verify that D is non-explosive and $\emptyset$ is an ergodic atom at which D regenerates, see Fig. 3.1.

We show next how X_t can be coupled to D_t forwards in time $t \in \mathbb{R}$. For $x \in \Omega$ and $\xi \in S \setminus x$, define

$$b(x, \xi) = \phi(x \cup \xi)/\phi(x) \tag{3.6}$$

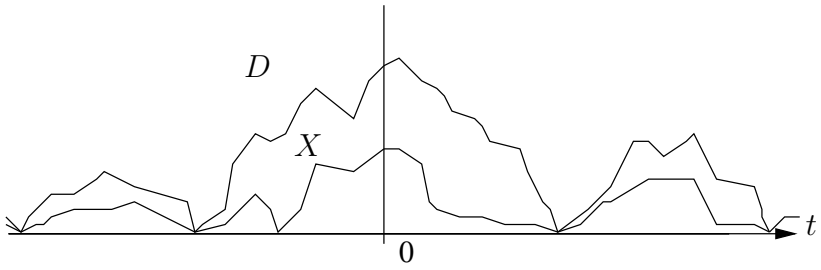

Figure 3.1: Upper curve: the dominating spatial birth-and-death process D; lower curve: the spatial birth-and-death process X. The horizontal axis is time and the vertical line corresponds to the state space Ω with $\emptyset$ placed at the bottom. Each time $D_t = \emptyset$, the jump process (D, X) regenerates.

(setting $0/0 = 0$). By (3.1), $b \leq 1$, and βb is a so-called Papangelou conditionally intensity (Kallenberg 1984). Consider a cycle of D given by $\{D_t : \tau_1 \leq t < \tau_2\}$ where τ_1 and τ_2 are two successive times at which D enters $\emptyset$, i.e. $D_{\tau_1-} \neq \emptyset$, $D_{\tau_1} = \emptyset$, and $\tau_2 = \inf\{t > \tau_1 : D_{t-} \neq \emptyset, D_t = \emptyset\}$ (with probability one, D enters $\emptyset$ infinite often, and $-\infty < \tau_1 < \tau_2 < \infty$). Then set $X_{\tau_1} = \emptyset$ and construct X_t forwards in time $t \in (\tau_1, \tau_2)$, according to the following rules:

$$\begin{aligned} D_t = D_{t-} &\quad\Rightarrow\quad X_t = X_{t-} \\ D_t = D_{t-} \cup \xi_t &\quad\Rightarrow\quad X_t = \begin{cases} X_{t-} \cup \xi_t & \text{if } R_t \leq b(X_{t-}, \xi_t) \\ X_{t-} & \text{otherwise} \end{cases} \\ D_t = D_{t-} \setminus \eta_t &\quad\Rightarrow\quad X_t = X_{t-} \setminus \eta_t. \end{aligned}$$

Using this coupling construction for all cycles of D, (3.5) is obviously satisfied.

It follows immediately from the coupling construction that X is a spatial birth-and-death process with birth rate βb and death rate 1. As ϕ satisfies the detailed balance condition $\phi(x)b(x, \xi) = \phi(x \cup \xi)$, we obtain that X is reversible with invariant (unnormalized) density ϕ. Hence, since (D, X) is time-stationary, X_t follows ϕ for any fixed time $t \in \mathbb{R}$.

In the case where $\phi(x) = \alpha^{n(x)}$ with $0 \leq \alpha \leq 1$, we have that (D, X) is reversible, X and $\{D_t \setminus X_t : t \in \mathbb{R}\}$ are independent spatial birth-and-death processes, and for any fixed time $t \in \mathbb{R}$, X_t and $D_t \setminus X_t$ are independent Poisson processes with intensity measures $\alpha\beta\lambda$ and $(1 - \alpha)\beta\lambda$, respectively. However, it is easily checked that (D, X) is in general not reversible, and apart from the Poisson case above, it seems complicated to obtain a closed form expression for the equilibrium distribution of (D, X).

3.4 The simple dominated CFTP algorithm

A jump in D happens when $D_t \neq D_{t-}$, in which case t is called a jump time. In order to generate a simulation of $X_0 \sim \phi$ we need only to consider the jump

chain (or embedded Markov chain) of $\{D_t : t < 0\}$, its associated marks for forwards births, and the states of X when $\{D_t : t < 0\}$ jumps. This is described in detail below.

Let $\ldots, Z_{-2}, Z_{-1}, Z_0$ denote the jump chain of $\{D_t : t < 0\}$ so that $Z_0 = D_0 \sim \nu$. This can be generated backwards in time together with the associated marks for forwards births as follows. For $i = 0, -1, -2, \ldots$,

- with probability $\beta/(\beta + n(Z_i))$ make a backwards birth:
  draw $\eta_i \sim \lambda$ and set $Z_{i-1} = Z_i \cup \eta_i$;
- else make a backwards death:
  draw randomly uniformly $\xi_i \in Z_i$, set $Z_{i-1} = Z_i \setminus \xi_i$, and generate the associated mark $R_i \sim \text{Uniform}[0, 1]$ for the forwards birth $Z_i = Z_{i-1} \cup \xi_i$.

Let $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$ and define

$$T_0 = \inf\{i \in \mathbb{N}_0 : Z_{-i} = \emptyset\}.$$

Furthermore, define recursively $Y_{-T_0}, \ldots, Y_0$, setting $Y_{-T_0} = \emptyset$ and using the rules

$$Z_i = Z_{i-1} \cup \xi_i \quad \Rightarrow \quad Y_i = \begin{cases} Y_{i-1} \cup \xi_i & \text{if } R_i \le b(Y_{i-1}, \xi_i) \\ Y_{i-1} & \text{otherwise} \end{cases} \tag{3.7}$$

$$Z_i = Z_{i-1} \backslash \eta_i \quad \Rightarrow \quad Y_i = Y_{i-1} \backslash \eta_i \tag{3.8}$$

for $i = -T_0 + 1, \ldots, 0$. Let $\cdots < \tau_{-2} < \tau_{-1} < \tau_0$ denote the jump times of D before time 0. Then $(X_{\tau_{-T_0}}, \ldots, X_{\tau_0})$ and $(Y_{-T_0}, \ldots, Y_0)$ follow the same distribution. Especially, $Y_0 \sim \phi$, since $X_{\tau_0} = X_0$ almost surely. This suggests the following perfect sampler.

The simple dominated CFTP algorithm

1. Generate backwards $Z_0, \ldots, Z_{-T_0}$, starting with $Z_0 \sim \nu$, and generate the associated marks R_i for forwards births $Z_i = Z_{i-1} \cup \xi_i$;
2. set $Y_{-T_0} = \emptyset$ and construct $Y_{-T_0+1}, \ldots, Y_0$ as in (3.7)–(3.8);
3. return $Y_0 \sim \phi$; see Fig. 3.2.

Proposition: The mean number of steps involved in the backwards construction of the simple dominated CFTP algorithm is bounded from below by

$$\mathbb{E}T_0 \ge \exp(\beta) - 1/2. \tag{3.9}$$

Proof: Let $Z_1, Z_2, \ldots$ denote the jump chain of $\{D_t : t > 0\}$. Set $M_i = n(Z_i)$ for $i \in \mathbb{Z}$, $T_0^- = T_0$, $T_0^+ = \inf\{i \in \mathbb{N}_0 : M_i = 0\}$, $L = \inf\{i \in \mathbb{N} : M_{i+T_0^+} = 0\}$, and $L_0 = T_0^- + T_0^+$ if $M_0 \neq 0$ and $L_0 = L$ otherwise. By time-stationarity,

$$\mathbb{E}L_0 \ge \mathbb{E}L = 1/\pi_0, \tag{3.10}$$

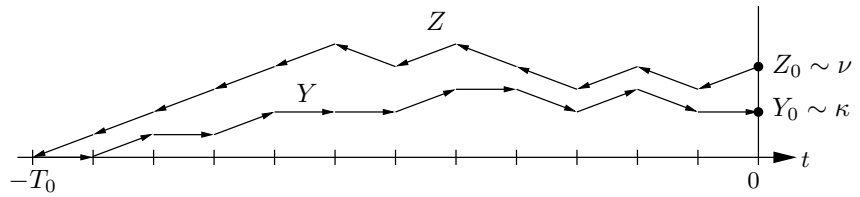

Figure 3.2: Illustration of the simple dominated CFTP algorithm.

where π denotes the invariant probability density function of M. By reversibility, T_0^- and T_0^+ are identically distributed, so

$$2\mathbb{E}T_0 = \mathbb{E}\big(T_0^- + T_0^+\big) = \mathbb{E}\big(L_0 \mathbf{1}[M_0 \neq 0]\big). \tag{3.11}$$

Further,

$$\mathbb{E}\big(L_0 \mathbf{1}[M_0 = 0]\big) = \pi_0 \mathbb{E}(L_0 | M_0 = 0) = \pi_0 \mathbb{E}L = 1. \tag{3.12}$$

Combining (3.10)–(3.12) we obtain that

$$\mathbb{E}T_0 \geq (1/\pi_0 - 1)/2.$$

Finally, by detailed balance of M,

$$\pi_i \beta/(\beta + i) = \pi_{i+1}(i+1)/(\beta + i + 1)$$

so by induction

$$\pi_{i+1} = \pi_0 \beta^i (\beta + i + 1)/(i+1)!, \quad i \in \mathbb{N}_0,$$

whereby $1/\pi_0 = 2\exp(\beta)$, and so (3.9) follows.

Remark: Since $\mathbb{E}T_0$ is at least exponentially growing in β, the simple dominated CFTP algorithm is infeasible for real applications of interest. For instance, if $\beta = 100$, then $\mathbb{E}T_0 \geq e^{100} - 1/2 \approx 2.7 \times 10^{43}$.

3.5 Upper and lower processes

A much faster perfect simulation algorithm is given in Kendall & Møller (2000), using upper and lower processes $U^j = \{U_j^j, \ldots, U_0^j\}$ and $L^j = \{L_j^j, \ldots, L_0^j\}$ which are started at times $j = 0, -1, -2, \ldots$. For each j, the upper and lower processes are constructed as follows. Initially set $U_j^j = Z_j$ and $L_j^j = \emptyset$. For

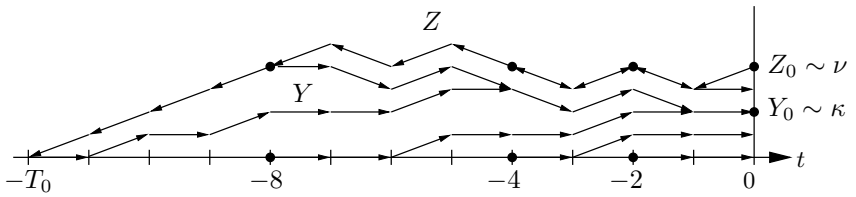

Figure 3.3: Illustration of sandwiching, funnelling, and coalescence properties for $T_0 = 12$ and $j = -2, -4, -8$ in (3.17)–(3.19).

$i = j+1, \ldots, 0$, if $z = Z_{i-1}$, $u = U_{i-1}^j$, and $l = L_{i-1}^j$, use the rules

$$Z_i = z \backslash \eta_i \quad \Rightarrow \quad U_i^j = u \backslash \eta_i \quad \text{and} \quad L_i^j = l \backslash \eta_i, \tag{3.13}$$

$$Z_i = z \cup \xi_i \quad \Rightarrow \quad U_i^j = \begin{cases} u \cup \xi_i & \text{if } R_i \le \alpha_{\max}(u, l, \xi_i) \\ u & \text{otherwise} \end{cases}$$

$$\text{and} \quad L_i^j = \begin{cases} l \cup \xi_i & \text{if } R_i \le \alpha_{\min}(u, l, \xi_i) \\ l & \text{otherwise,} \end{cases} \tag{3.14}$$

where

$$\alpha_{\max}(u, l, \xi) = \max\{b(x, \xi) : l \subseteq x \subseteq u\} \tag{3.15}$$

and

$$\alpha_{\min}(u, l, \xi) = \min\{b(x, \xi) : l \subseteq x \subseteq u\}. \tag{3.16}$$

Notice that $U^j, L^j, U^{j-1}, L^{j-1}, \ldots$ are coupled by the same R_i, ξ_i, η_i for $i > j$.

The construction in (3.13)–(3.16) ensures the sandwiching property

$$L_i^j \subseteq Y_i \subseteq U_i^j \subseteq Z_i, \quad j \le i \le 0, \tag{3.17}$$

the funneling property

$$L_i^{j'} \subseteq L_i^j \subseteq U_i^j \subseteq U_i^{j'}, \quad j \le j' \le i \le 0, \tag{3.18}$$

and the coalescence property

$$L_i^j = U_i^j \quad \Rightarrow \quad L_{i'}^j = U_{i'}^j \quad \text{for } j \le i \le i' \le 0, \tag{3.19}$$

see Fig. 3.3. The sandwiching property explains why the U^j and L^j are called upper and lower processes: they bound the "target process" Y. The definitions (3.15)–(3.16) seem natural as they provide the minimal upper and maximal lower processes so that (3.17) is satisfied for all possible realizations of the marks R_i. By (3.17) and (3.19), once a pair of upper and lower processes have coalesced, they stay in coalescence, and at time 0 they are equal to $Y_0 \sim \phi$.

The time

$$T = \inf\{j \in \mathbb{N}_0 : U_0^{-j} = L_0^{-j}\}$$

is called the true coalescence time for upper and lower processes. The funneling property (3.18) suggests that instead of searching for T it may be advantageous to search for a larger coalescence time. Therefore, consider any sequence of (possibly random) integers $\dots j_2 < j_1 < 0$ such that $\lim_{k\to\infty} j_k = -\infty$, and let

$$T_{\{j_k\}} = \inf\{-j_k : U_0^{j_k} = L_0^{j_k}\}$$

be the first time that $U_0^j = L_0^j = Y_0$ when pairs of upper and lower processes are started at times $j = j_1, j_2, \dots$. Further, let

$$T_{\min} = \inf\{i \in \mathbb{N}_0 : Z_{-i} \cap Z_0 = \emptyset\}$$

be the time just before the first point in Z_0 is born. For $-T_{\min} < j \le 0$, we have that $U_0^j \supseteq Z_j \cap Z_0 \neq \emptyset$ and $L_0^j \cap Z_j = \emptyset$, so clearly

$$T_{\min} \le T \le T_{\{j_k\}} \le T_0. \tag{3.20}$$

For efficiency reasons a doubling scheme is usually used (Propp & Wilson 1996, Wilson 2000b), i.e. $j_k = -2^{k-1}n$, where $n \in \mathbb{N}$ is chosen by the user; then we write T_n for $T_{\{j_k\}}$. Typically in applications $T_n \ll T_0$, cf. Section 3.7. Taking (3.20) into account, we propose to replace n by $T_{\min}$ in the doubling scheme; then we write T_* for $T_{\{j_k\}}$. See also the empirical results in Section 3.7.

Given a sequence of (possibly random) integers $\dots j_2 < j_1 < 0$ such that $\lim_{k\to\infty} j_k = -\infty$, we have the following perfect sampler, where we set $j_0 = 0$.

The dominated CFTP algorithm based on upper and lower processes

1. Generate $Z_0 \sim \nu$;
2. repeat the following steps 3.–4. for $k = 1, 2, \dots$ until $U_0^{j_k} = L_0^{j_k}$;
3. generate backwards $Z_{j_{k-1}-1}, \dots, Z_{j_k}$ and generate the associated marks $R_i \sim \text{Uniform}[0,1]$ each time $Z_i \setminus Z_{i-1} \neq \emptyset,\ j_k < i \le j_{k-1}$;
4. generate forwards $(U_{j_k}^{j_k}, L_{j_k}^{j_k}), \dots, (U_0^{j_k}, L_0^{j_k})$ as in (3.13)–(3.14);
5. return $U_0^{-T_{\{j_k\}}} \sim \phi$.

The calculation of $\alpha_{\max}$ and $\alpha_{\min}$ is particularly simple in the following cases. A point process is attractive if

$$b(x,\xi) \le b(y,\xi) \qquad \text{whenever} \qquad x \subseteq y, \xi \notin y, \tag{3.21}$$

and repulsive if

$$b(x,\xi) \ge b(y,\xi) \qquad \text{whenever} \qquad x \subseteq y, \xi \notin y. \tag{3.22}$$

For the Strauss and area interaction point processes (Examples 1 and 2), either (3.21) or (3.22) is satisfied. In the attractive case (3.21), $\alpha_{\max}(u, l, \xi) =$

$b(u,\xi)$ and $\alpha_{\min}(u,l,\xi) = b(l,\xi)$, while in the repulsive case (3.22), $\alpha_{\max}(u,l,\xi) = b(l,\xi)$ and $\alpha_{\min}(u,l,\xi) = b(u,\xi)$. Note that it is only in the attractive case that U^j and L^j are individual Markov chains.

It other situations it may be quite time consuming to calculate $\alpha_{\max}$ and $\alpha_{\min}$ by (3.15) and (3.16). For instance, if $b(x,\xi) = b_a(x,\xi)b_r(x,\xi)$ factorizes into two terms, where $b_a(x,\xi) \leq K_a$ is increasing in x, $b_r(x,\xi) \leq K_r$ is decreasing in x, and $K_aK_r \leq 1$, it may be convenient to redefine $\alpha_{\max}$ and $\alpha_{\min}$ by

$$\alpha_{\max}(u,l,\xi) = b_a(u,\xi)b_r(l,\xi) \quad \text{and} \quad \alpha_{\min}(u,l,\xi) = b_a(l,\xi)b_r(u,\xi).$$

Since (3.17)–(3.19) are satisfied with this choice of $\alpha_{\max}$ and $\alpha_{\min}$, the algorithm still works.

Example 1 (continued): Perfect simulations of different Strauss processes with $\gamma = 1$ (the Poisson case), $\gamma = 0.5$, and $\gamma = 0$ are shown in Fig. 3.4, using the same dominating process (and associated marks) in all three cases. Due to the thinning procedure in the algorithm, the point pattern with $\gamma = 1$ contains the two others. The point pattern with $\gamma = 0$ does not contain the point pattern with $\gamma = 0.5$, because the Strauss process is repulsive.

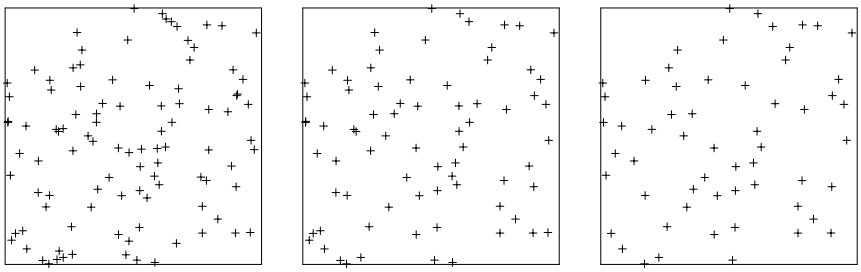

Figure 3.4: Simulation of a Strauss process on $S = [0,1]^2$, when $\beta = 100$, $R = 0.05$, and $\gamma = 1, 0.5, 0$ (from left to right).

3.6 Clan of ancestors

In this section we consider an alternative algorithm due to Fernández et al. (2000). For simplicity we assume that S is a metric space and ϕ has finite range of interaction, i.e. there exists an $R < \infty$ such that for any $x \in \Omega$ and $\xi \in S \backslash x$, $b(x,\xi) = b(x \cap \text{ball}(\xi,R),\xi)$, where $\text{ball}(\xi,R)$ denotes the ball with center ξ and radius R. This is fulfilled in Examples 1 and 2.

In order to understand the following definitions it may be useful to consider Fig. 3.5 and to keep in mind how the simple dominated CFTP algorithm (Section 3.4) works. For $\xi \in \cup_{i \leq 0} Z_i$, let $I(\xi)$ be the time at which ξ was born, i.e. $I(\xi) = i$ if $\xi = \xi_i$ in (3.7). We call

$$\text{an}^1(\xi) = Z_{I(\xi)-1} \cap \text{ball}(\xi,R)$$

the first generation of ancestors of ξ, define recursively the jth generation of ancestors of ξ by

$$\mathrm{an}^j(\xi) = \cup_{\eta \in \mathrm{an}^{j-1}(\xi)} \mathrm{an}^1(\eta), \quad j = 2, 3, \ldots,$$

and call $\mathrm{an}(\xi) = \cup_{j \in \mathbb{N}} \mathrm{an}^j(\xi)$ the ancestors of ξ. If $I(\xi) = i$, then $Y_{i-1} \cap \mathrm{ball}(\xi, R) \subseteq \mathrm{an}(\xi)$, so the ancestors of $\xi = \xi_i$ are the only points in Z which are needed in (3.7) in order to determine whether or not $\xi_i \in Y_i$. Hence Y_0 depends only on Z through $Z_C = C(Z_0) \cup Z_0$, where

$$C(Z_0) = \cup_{\xi \in Z_0} \mathrm{an}(\xi)$$

is called the *clan of ancestors* of Z_0. Finally, let

$$T_C = \inf\{i \in \mathbb{N}_0 : Z_{-i} \cap Z_C = \emptyset\}$$

specify the time interval in which the points in Z_C are living. Then $T_C \leq T_0$, and Y_0 is unaffected if we set $Y_{-T_C} = \emptyset$ and generate Y_i forwards in time $i \geq -T_C$ as usual, but considering only the transitions in $Z_{-T_C} \cap Z_C, \ldots, Z_0 \cap Z_C$.

The dominated CFTP algorithm based on the clan of ancestors

1. Generate backwards $Z_0, \ldots, Z_{-T_C}$, i.e. starting with $Z_0 \sim \nu$;
2. set $Y_{-T_C} = \emptyset$ and generate forwards $Y_{-T_C+1}, \ldots, Y_0$ as in (3.7)– (3.8), but so that $Y_{i+1} = Y_i$ is unchanged whenever $Z_{i+1} \cap Z_C = Z_i \cap Z_C$ is unchanged;
3. return $Y_0 \sim \phi$.

It is not hard to see that $T \leq T_C$, so

$$T \leq T_C \leq T_0. \tag{3.23}$$

Note that T_C depends only on b through R, and no monotonicity properties such as (3.21) and (3.22) are required. The algorithm can easily be modified to perfect Metropolis-Hastings simulation of locally stable point processes (Kendall & Møller 2000), and to perfect Gibbs sampling of the Widom & Rowlinson (1970) model (Häggström et al. 1999) and related models (Georgii 2000). The case of the Widom-Rowlinson model turns out to be particular simple.

3.7 Empirical findings

In this section we present some empirical findings for the dominated CFTP algorithm based on upper and lower processes (Section 3.5) and the clan algorithm (Section 3.6), respectively. The algorithms are applied on a Strauss process defined on the unit square with $\beta = 100$ and $R > 0$ (Example 1). Note that as the interaction parameter γ increases, the interaction/repulsion between the points

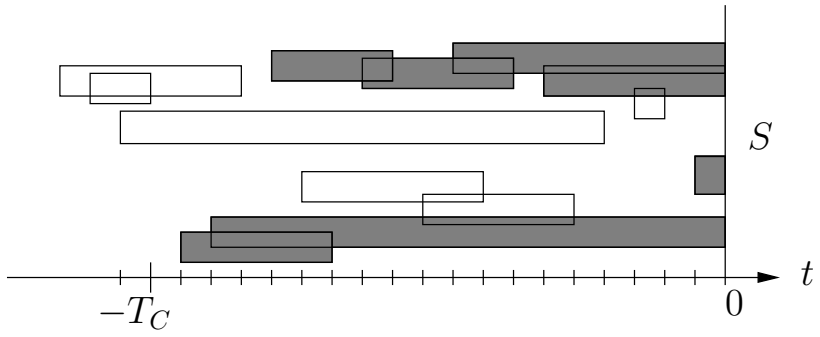

Figure 3.5: Example of a clan of ancestors when S is a line segment. The points in D agree with the midpoints of the vertical edges of the rectangles. Each horizontal edge of a rectangle shows the life time of the corresponding point in D. The vertical edges are all of length R. Shaded rectangles represent members of the clan.

in the Strauss process decreases. Below we consider three values of γ: $\gamma = 0$ (a so-called hard core process), $\gamma = 0.5$, and $\gamma = 1$ (a Poisson process on the unit square with rate $\beta = 100$).

First we consider the algorithm based on upper and lower processes, using the doubling scheme with either $n = 1$ or n replaced by $T_{\min}$. Recall that T_1 and T_* denote the corresponding coalescence times, cf. Section 3.5. The number of steps involved in the backward-forwards construction in the two cases are given by

$$N_1 \equiv T_1 + (1 + 2 + 4 + \ldots + T_1) = 3T_1 - 1$$

and

$$N_{\min} \equiv T_* + (T_{\min} + 2T_{\min} + 4T_{\min} + \ldots + T_*) = 3T_* - T_{\min},$$

respectively. It makes sense to compare N_1 and $N_{\min}$ because the "basic algorithm" is the same in the two cases.

The left plot in Fig. 3.6 shows how the means $\mathbb{E}N_1$ and $\mathbb{E}N_{\min}$ depend on $R > 0$ when $\gamma = 0$. Each mean is estimated by the empirical average based on 500 independent runs of the algorithm. For all values of R in the plot, $\mathbb{E}N_{\min} < \mathbb{E}N_1$, but the difference decreases as R increases. Based on this and other results (not shown here) we prefer to replace n by $T_{\min}$ in the doubling scheme.

Next we compare the dominated CFTP algorithm based on upper and lower processes and the clan algorithm. As these algorithms are not immediately comparable, there is little sense in comparing the number of steps involved in the backwards-forwards construction in the two algorithms. Instead we just consider the means $\mathbb{E}T_*$ and $\mathbb{E}T_C$ for $R > 0$ and either $\gamma = 0$ or $\gamma = 0.5$,

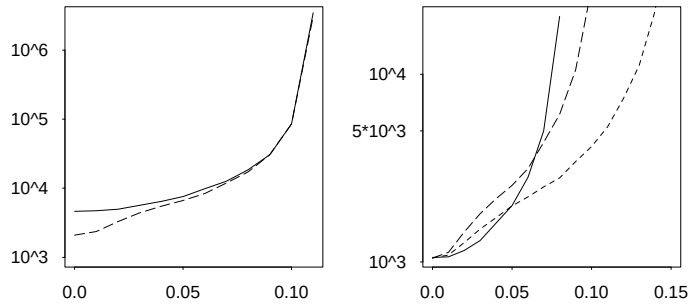

Figure 3.6: Various mean values related to the CFTP algorithms, where each mean is estimated from 500 independent runs. Left plot: $\mathbb{E}N_1$ (full line) and $\mathbb{E}N_{\min}$ (dotted line) versus R. Right plot: $\mathbb{E}T_C$ (full line) and $\mathbb{E}T_*$ when $\gamma = 0$ (upper dotted line) and $\gamma = 0.5$ (lower dotted line) versus R.

though this is of course not telling the whole story about which algorithm is the fastest.

The right plot in Fig. 3.6 shows $\mathbb{E}T_*$ and $\mathbb{E}T_C$ versus R when $\gamma = 0$ and $\gamma = 0.5$, respectively. Note that T_C does not depend on γ, and each mean in the plot is estimated by the empirical average based on 500 independent runs of the algorithm. All means in the plot are much smaller than $\mathbb{E}T_0 \geq e^{100} - 1/2$, cf. (3.9). As expected the means agree as R tends to 0, and $\mathbb{E}T_*$ decreases as γ increases. For both $\gamma = 0$ and $\gamma = 0.5$, it is only for rather small values of γ that $\mathbb{E}T_*$ is larger than $\mathbb{E}T_C$. The picture changes as γ tends to 1, since in the limit T_* agrees with $T_{\min}$ which is smaller than T_C, cf. (3.20) and (3.23). Furthermore, as R increases, $\mathbb{E}T_*$ becomes much smaller than $\mathbb{E}T_C$.

We have also investigated empirically how $\mathbb{E}T_C$ and $\mathbb{E}T_*$ depend on β, and obtained similar conclusions as above. Further empirical results for the Strauss process and other locally stable point processes can be found in Berthelsen & Møller (2001a).

Finally, all things considered our conclusion is that the dominated CFTP algorithm based on upper and lower processes using the doubling scheme with n replaced by $T_{\min}$ seems to be the best choice.

3.8 Perfect simulation of infinite point processes

Often one considers point processes with infinitely many points contained in an "infinite volume" such as $\mathbb{R}^d$. In order to avoid edge-effects, a perfect sample within a bounded region may be achieved by extending simulations both backwards in time and space (Kendall 1997, Fernández et al. 2000, Georgii 2000). This is sometimes possible, for example if b is sufficiently close to 1 and the inter-

action radius R is sufficiently small. The constructions in the abovementioned papers are rather straightforward, but particularly the algorithm in Fernández et al. (2000) allows a detailed mathematical analysis. Such coupling constructions may be of great theoretical interest, but in our opinion they remain so far unpractical for applications of real interest.

Acknowledgment

This paper will be a part of KKB's PhD dissertation. JM was supported by the European Union's research network "Statistical and Computational Methods for the Analysis of Spatial Data, ERB-FMRX-CT96-0096", by the Centre for Mathematical Physics and Stochastics (MaPhySto), funded by a grant from the Danish National Research Foundation, and by the Danish Natural Science Research Council. JM also acknowledges the invitation and support for attending the 5th Brazilian School of Probability.

Bibliography

Baddeley, A. & Møller, J. (1989). Nearest-neighbour markov point processes and random sets, *Internat. Statist. Rev.* **2**: 89–121.

Baddeley, A. & van Lieshout, M. N. M. (1995). Area-interaction Point Processes, *Ann. Inst. Statist. Math.* **46**: 601–619.

Berthelsen, K. K. & Møller, J. (2001a). Perfect simulation and inference for spatial point processes. In preparation.

Berthelsen, K. K. & Møller, J. (2001b). Spatial jump processes and perfect simulation, *Technical report*, R-01-2008, Department of Mathematical Sciences, Aalborg University. Submitted.

Fernández, R., Ferrari, P. A. & Garcia, N. L. (2000). Perfect simulation for interacting point processes, loss networks and Ising models. Submitted to *Stoch. Process. Appl.*

Georgii, H.-O. (2000). Phase transitions and percolation in Gibbsian particle models, *in* K. R. Mecke & D. Stoyan (eds), *Statistical Physics and Spatial Statistics*, Vol. 554 of *Lecture Notes in Physics*, Springer, Berlin, pp. 267–294.

Geyer, C. J. (1999). Likelihood inference for spatial processes, *in* O. E. Barndorff-Nielsen, W. Kendall & M. van Lieshout (eds), *Stochastic Geometry, Likelihood and Computation*, Chapman & Hall, pp. 79–140.

Geyer, C. J. & Møller, J. (1994). Simulation procedures and likelihood inference for spatial point processes, *Scand. J. Statist.* **21**: 359–373.

Häggström, O., van Lieshout, M. N. M. & Møller, J. (1999). Characterization results and Markov chain Monte Carlo algorithms including exact simulation for some spatial point processes, *Bernoulli* **5**: 641–658.

Kallenberg, O. (1984). An informal guide to the theory of conditioning in point processes, *Int. Statist. Rev.* **52**: 151–164.

Kelly, F. P. & Ripley, B. D. (1976). A note on Strauss's model for clustering, *Biometrika* **63**: 357–360.

Kendall, W. S. (1997). Perfect simulation for spatial point processes, *Proc. ISI 51st session, Istanbul* **3**: 163–166.

Kendall, W. S. (1998). Perfect simulation for the area-interaction point process, *in* L. Accardi & C. Heyde (eds), *Probability Towards 2000*, Springer, New York, pp. 218–234.

Kendall, W. S. & Møller, J. (2000). Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes, *Adv. Appl. Prob.* **32**: 844–865.

Lieshout, M. N. M. van (2000). *Markov point processes and their applications*, Imperial College Press, London.

Møller, J. (1989). On the rate of convergence of spatial birth-and-death processes, *Ann. Inst. Statist. Math.* **41**: 565–581.

Møller, J. (2001). A review of perfect simulation in stochastic geometry, *in* C. C. H. I. V. Basawa & R. L. Taylor (eds), *Selected Proceedings of the Symposium on Inference for Stochastic Processes*, Vol. 37, IMS Lecture Notes & Monographs Series, pp. 333–355.

Preston, C. (1977). Spatial birth-and-death processes, *Bull. Int. Statist. Inst.* **46**: 371–391.

Propp, J. G. & Wilson, D. B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics, *Random Structures and Algorithms* **9**: 223–252.

Ruelle, D. (1969). *Statistical Mechanis: Rigorous Results*, W. A. Benjamin, Reading, Massachusetts.

Strauss, D. J. (1975). A model for clustering, *Biometrika* **62**: 467–475.

Widom, B. & Rowlinson, J. S. (1970). A new model for the study of liquid-vapor phase transitions, *J. Chem. Phys.* **52**: 1670–1684.

Wilson, D. B. (2000a). How to couple from the past using a read-once source of randomness, *Random Structures and Algorithms* **16**: 85–113.

Wilson, D. B. (2000b). Layered multishift coupling for use in perfect sampling algorithms (with a primer on CFTP), *in* N. Madras (ed.), *Monte Carlo Methods*, Vol. 26 of *Fields Institute Communications*, Amer. Stat. Soc., pp. 141–176.

4 Likelihood and non-parametric Bayesian MCMC inference for spatial point processes based on perfect simulation and path sampling

Kasper K. Berthelsen and Jesper Møller

Department of Mathematical Sciences, Aalborg University,
Fredrik Bajers Vej 7G, DK-9220 Aalborg, Denmark

Abstract. We consider the combination of path sampling and perfect simulation in the context of both likelihood inference and non-parametric Bayesian inference for pairwise interaction point processes. Several empirical results based on simulations and analysis of a dataset are presented, and the merits of using perfect simulation are discussed.

Keywords: Coupling from the past (CFTP), dominated CFTP, exact simulation, likelihood inference, Markov chain Monte Carlo, multiscale process, non-parametric Bayesian smoothing, path sampling, pairwise interaction point process, simulation-based inference, Strauss process.

4.1 Introduction

Since the seminal work by Propp & Wilson (1996) on perfect (or exact) simulation many new perfect sampling algorithms have been developed, particularly for spatial point processes. Meanwhile the possibilities of using perfect simulation for statistical inference have been less explored. A few papers (Lund & Thönnes 2000, van Lieshout & van Zwet 2001, Loizeaux & McKeague 2001, Møller 2001, McKeague & Loizeaux 2002) explore this in the context of spatial applications. This paper concerns simulation-based inference for spatial point process models using Markov chain Monte Carlo (MCMC) methods, including perfect simulation techniques.

We restrict attention to pairwise interaction point processes defined on a bounded planar region S by a density

$$f_{\beta,\varphi}(x) = \left[\beta^{n(x)} \prod_{\{\xi,\eta\}\subseteq x:\, \xi\neq\eta} \varphi(\|\xi-\eta\|)\right] \Big/ c_{\beta,\varphi} \tag{4.1}$$

with respect to the unit rate Poisson process on S. Here x denotes any finite

subset of S, $\beta > 0$ is a parameter, $n(x)$ is the cardinality of x, $\varphi \geq 0$ is an interaction function, $\|\cdot\|$ denotes Eucledian distance, and $c_{\beta,\varphi}$ is a normalising constant. Some restrictions on φ are needed to ensure integrability; we assume that $\varphi \leq 1$. In general a closed form expression of $c_{\beta,\varphi}$ is unknown. Sections 4.3 and 4.4 discuss the advantages and the computational limitations of using perfect simulation when the normalising constant is estimated by a useful technique called path sampling (Gelman & Meng 1998).

In many applications φ is of primary interest, and one sometimes conditions on $n(x)$. The reason for conditioning is that this conditional distribution does not depend on β, though it is unclear whether $n(x)$ in some sense is an ancillary statistic; see also the discussion in Gates & Westcott (1986), Ripley (1988), and Geyer & Møller (1994). Furthermore, there is at present no practical way of generating perfect simulations conditional on $n(x)$. We have therefore chosen not to fix $n(x)$.

The paper is organised as follows.

Section 4.2 provides a short description of perfect simulation based on the dominated coupling from past algorithm (dominated CFTP; Kendall 1998, Kendall & Møller 2000), and shows some empirical findings for the running time of the algorithm.

Section 4.3 discusses how dominated CFTP and path sampling can be combined to obtain a Monte Carlo approximation of the normalising constant when φ belongs to a parametric class of models of low dimension. For specificity we consider a Strauss process given by

$$\varphi(r) = \gamma^{\mathbf{1}[r \leq R]} \tag{4.2}$$

where $0 < \gamma \leq 1$ is an interaction parameter, $R > 0$ is the range of interaction, and $\mathbf{1}[\cdot]$ denotes the indicator function (Strauss 1975, Kelly & Ripley 1976). Often in the literature R is assumed to be known, see e.g. Ripley (1989) and Baddeley & Turner (2000), but we treat all three parameters β, γ, R as unknown parameters. In Example 11, Section 4.3, we estimate the likelihood surface and discuss how approximate MLE and likelihood ratio test statistics perform.

Section 4.4 is the main section. It concerns non-parametric Bayesian analysis for the pairwise interaction point process (4.1) when φ is approximated by a step function,

$$\varphi(r) = \sum_{i=1}^{p} \gamma_i \mathbf{1}[r_{i-1} < r \leq r_i] + \mathbf{1}[r > r_p] \tag{4.3}$$

where $p \in \mathbb{N}_0$ is the number of change-points $r_1, \ldots, r_p$ with $r_0 = 0 < r_1 < \ldots < r_p$, and where $0 < \gamma_1 < \ldots < \gamma_p < 1$. This is an obvious extension of (4.2) called the multiscale process (Penttinen 1984). We impose a prior for (β, ψ) where $\psi = \{(r_1, \gamma_1), \ldots, (r_p, \gamma_p)\}$ is viewed as a marked point process, and discuss how a fully Bayesian MCMC analysis can be performed when the unknown normalising constant from the likelihood term is approximated by path sampling. A similar situation is considered in Heikkinen & Penttinen (1999), but they condition on $n(x)$ in the likelihood, and fix p and $(r_1, \ldots, r_p)$ in the prior.

They develop an MCMC root-finding algorithm for the maximum a posteriori estimate of $(\gamma_1, \ldots, \gamma_p)$, but mention at the end of their paper the advantage of estimating φ by the Monte Carlo posterior mean, which is usually a smooth curve (Arjas & Gasbarra 1994). Among other things we consider such estimates.

Several empirical results based on simulations and analysis of a dataset are discussed in Sections 4.2–4.4. Finally, Section 4.5 contains some concluding remarks, including a discussion of the merits and limitations of perfect simulation for simulation-based inference.

4.2 Perfect simulation using dominated CFTP

Different kinds of perfect samplers have been developed for spatial point processes, cf. the surveys in Møller (2001), Berthelsen & Møller (2002), and the references therein. In general dominated CFTP is the most applicable algorithm for pairwise interaction point processes; for extensions to locally stable point processes, see Kendall & Møller (2000) and Berthelsen & Møller (2002).

The algorithm uses a so-called dominating process D_i, $i = 0, -1, -2, \ldots$, of finite point configurations contained in S. This is a Markov chain, which is generated backwards in time as follows. Let ν_β denote the homogeneous Poisson point process on S with rate $\beta > 0$, and let $\emptyset$ denote the empty point configuration. Then $D_0 \sim \nu_\beta$, and for $i = 0, -1, -2, \ldots$,

- with probability $\beta/(\beta + n(D_i))$ we make a backwards birth: we generate a point η_i which is uniformly distributed on S, and set $D_{i-1} = D_i \cup \{\eta_i\}$;
- else we make a backwards death: we draw randomly uniformly a point $\xi_i \in D_i$, and set $D_{i-1} = D_i \setminus \{\xi_i\}$ (if $D_i = \emptyset$ we set $\{\xi_i\} = \emptyset$); for later use we also draw a "mark" $M_i \sim \text{Uniform}[0, 1]$.

The D_i and (in case of a backwards death) their marks M_i are easy to generate. They are used for generating so-called upper and lower processes $U^j = \{U_j^j, \ldots, U_0^j\}$ and $L^j = \{L_j^j, \ldots, L_0^j\}$, which are started at times $j = 0, -1, -2, \ldots$, and which are generated forwards in time until time 0 as follows. Initially set $U_j^j = D_j$ and $L_j^j = \emptyset$. For $i = j + 1, \ldots, 0$,

$$D_i = D_{i-1} \backslash \{\eta_i\} \quad \Rightarrow \quad U_i^j = U_{i-1}^j \backslash \{\eta_i\} \quad \text{and} \quad L_i^j = L_{i-1}^j \backslash \{\eta_i\}, \tag{4.4}$$

and

$$D_i = D_{i-1} \cup \{\xi_i\} \quad \Rightarrow \quad U_i^j = \begin{cases} U_{i-1}^j \cup \{\xi_i\} & \text{if } M_i \le \prod_{\eta \in L_{i-1}^j} \varphi(\|\xi_i - \eta\|) \\ U_{i-1}^j & \text{otherwise} \end{cases} \tag{4.5}$$

$$\text{and} \quad L_i^j = \begin{cases} L_{i-1}^j \cup \{\xi_i\} & \text{if } M_i \le \prod_{\eta \in U_{i-1}^j} \varphi(\|\xi_i - \eta\|) \\ L_{i-1}^j & \text{otherwise.} \end{cases} \tag{4.6}$$

Note that $(U^j, L^j), (U^{j-1}, L^{j-1}), \ldots$ are coupled by the same η_i, ξ_i, M_i for $i > j$.

Berthelsen & Møller (2002) propose to use a doubling scheme given by $j_k = -2^k T_{\min},\ k = 0, 1, 2, \ldots$, where

$$T_{\min} = \begin{cases} \inf\{-i : D_i \cap D_0 \neq \emptyset, D_{i-1} \cap D_0 = \emptyset\} & \text{if } D_0 \neq \emptyset \\ 0 & \text{otherwise} \end{cases}$$

is the first time a point in D_0 is born (when considering D forwards in time). We use the same scheme and let

$$T = \inf\{-j_k : U_0^{j_k} = L_0^{j_k}\}$$

denote the first time we have coalescence at time 0 when the upper and lower processes are started at times $j_0, j_1, j_2, \ldots$. We have that $U_0^{-T} \sim f_{\beta,\varphi}$.

To summarise we use the following dominated CFTP algorithm where we set $j_{-1} = 0$:

1. generate $D_0 \sim \nu_\beta$;
2. repeat the following steps 3–4 for $k = 0, 1, 2, \ldots$ until $U_0^{j_k} = L_0^{j_k}$;
3. generate backwards $D_{j_{k-1}-1}, \ldots, D_{j_k}$ and generate the associated marks $M_i \sim \text{Uniform}[0, 1]$ each time $D_i \setminus D_{i-1} \neq \emptyset,\ j_k < i \leq j_{k-1}$;
4. generate forwards $(U_{j_k}^{j_k}, L_{j_k}^{j_k}), \ldots, (U_0^{j_k}, L_0^{j_k})$ as in (4.4)–(4.6);
5. return $U_0^{-T} \sim f_{\beta,\varphi}$.

The computer time of the algorithm depends of course on how efficient the implementation is. It is advantageous to exploit the fact that the products in (4.5) and (4.6) depend only on local information. Specifically, if S is rectangular and $R = \sup\{r > 0 : \varphi(r) < 1\}$ is the range of interaction, we use a subdivision $S = \cup_k C_k$ of rectangular cells of side lengths $\geq R$, and exploit the fact that if $\xi \in C_k$, then $\prod_{\eta \in x} \varphi(\|\xi - \eta\|)$ depends only on the points from x falling in C_k and in the neighbouring cells to C_k.

Example 9: The dominated CFTP algorithm for the Strauss process (4.2) gets slower and slower as the interaction parameter γ decreases. Fig. 4.1 shows the mean of T in the limiting case of (4.2) as $\gamma \to 0$, i.e. a hard core point process given by $\varphi(r) = \mathbf{1}[r > R]$. For fixed $R = 0.1$, the figure indicates the existence of a "critical value" $\beta_{\text{crit}} \approx 100$, where $\log \mathbb{E}T$ is nearly linear for $\beta < \beta_{\text{crit}}$, and $\log \mathbb{E}T$ grows faster and faster for $\beta > \beta_{\text{crit}}$. For fixed $\beta = 100$, $\log \mathbb{E}T$ depends on R in a similar way but with respect to a critical value $R_{\text{crit}} \approx 0.1$.

4.3 Perfect simulation and likelihood inference for the Strauss process

In this section we consider a Strauss process with density

$$f_\theta(x) = \beta^{n(x)} \gamma^{s_R(x)} / c_\theta \tag{4.7}$$

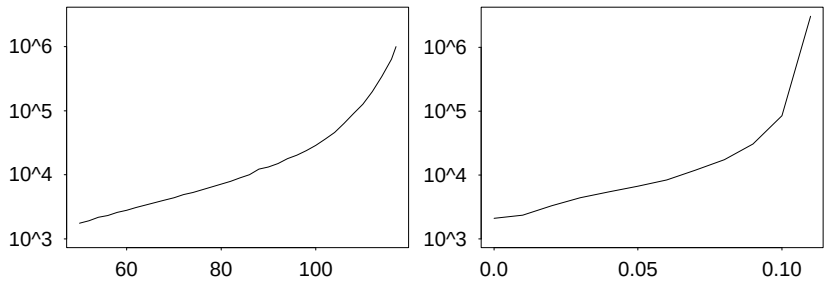

Figure 4.1: Mean coalescence times on a log scale for a hard core process on $S = [0,1]^2$. Left plot: $\mathbb{E}T$ versus β when $R = 0.1$. Right plot: $\mathbb{E}T$ versus R when $\beta = 100$.

where $\theta = (\beta, \gamma, R) \in (0,\infty) \times (0,1] \times (0,\infty)$ and

$$s_R(x) = \sum_{\{\xi,\eta\} \subseteq x} \mathbf{1}[0 < \|\xi - \eta\| \leq R]$$

is the number of R-close pairs of points in x.

4.3.1 Path sampling

The normalising constant of the Strauss process is unknown except for the Poisson process in which case $\gamma = 1$ and

$$c_\theta = \exp(\beta). \tag{4.8}$$

We discuss below how path sampling can be combined with perfect simulation for estimating ratios of normalising constants.

We use a simple version of path sampling; other versions are discussed in Gelman & Meng (1998); see also Section 4.5. We start by letting (β, R) be fixed, and notice the following identity: for $\theta_i = (\beta, \gamma_i, R),\ i = 1, 2$, with $0 < \gamma_1 < \gamma_2 \leq 1$,

$$\ln(c_{\theta_2}/c_{\theta_1}) = \int_{\gamma_1}^{\gamma_2} \mathbb{E}_\theta s_R/\gamma \, \mathrm{d}\gamma \tag{4.9}$$

where $\theta = (\beta, \gamma, R)$ varies with γ in the integral, and where $\mathbb{E}_\theta$ denotes expectation with respect to (4.7). Taking $\gamma_2 = 1$ we know c_{θ_2} by (4.8). Let $\gamma^{(j)} = \gamma_1 + j\delta,\ j = 0, \ldots, k$, be a grid of γ-values where $k \geq 1$ and $\delta = (\gamma_2 - \gamma_1)/k$. Estimating $\mathbb{E}_\theta s_R$ for $\theta = \theta^{(j)} = (\beta, \gamma^{(j)}, R),\ j = 0, \ldots, k-1$, as described below, we approximate the integral in (4.9) by a Riemann sum,

$$\ln(c_{\theta_2}/c_{\theta_1}) \approx \delta \left[\mathbb{E}_{\theta^{(0)}} s_R/(2\gamma^{(0)}) + \mathbb{E}_{\theta^{(k)}} s_R/(2\gamma^{(k)}) + \sum_{j=1}^{k-1} \mathbb{E}_{\theta^{(j)}} s_R/\gamma^{(j)} \right].$$

Similarly, for $j = 0, \ldots, k-1$,

$$\ln(c_{\theta_2}/c_{\theta^{(j)}}) \approx \delta\left[\mathbb{E}_{\theta^{(j)}} s_R/(2\gamma^{(j)}) + \mathbb{E}_{\theta^{(k)}} s_R/(2\gamma^{(k)}) + \sum_{l=j+1}^{k-1} \mathbb{E}_{\theta^{(l)}} s_R/\gamma^{(l)}\right]. \tag{4.10}$$

Thereby we obtain estimates of $c_{\theta^{(j)}}$, $j = 0, \ldots, k-1$. Finally, combining the grid of γ-values with a grid of (β, R)-values, all normalising constants c_θ can first be estimated for θ in the corresponding 3D-grid, and next by interpolation and extrapolation over a region of the parameter space.

For a given value of θ, we estimate $\mathbb{E}_\theta s_R$ by

$$\sum_{i=0}^{m} s_R(X_i)/(m+1) \tag{4.11}$$

where $X_0 \sim f_\theta$ is generated by the dominated CFTP algorithm, and where $X_1, \ldots, X_m$ is a sample of length $m \geq 0$ obtained by the Metropolis-Hastings algorithm in Geyer & Møller (1994). Then the estimate (4.11) is unbiased. See also the discussion in Section 4.5.

It remains to specify k and m. For (4.10) to be a "good" approximation we need a sufficiently large k. As k increases, the precision of the unbiased estimator (4.11) becomes less and less important. We can even take $m = 0$ for sufficiently large values of k. On the other hand, the perfect simulation of X_0 is computationally more demanding than a Metropolis-Hastings update $X_i \to X_{i+1}$. The "right" choice of (k, m) is context-dependent and requires some experimentation. This is illustrated in Example 10 below.

4.3.2 Empirical results

In the following two examples, $S = [0,1]^2$ is the unit square.

Example 10: In Fig. 4.2 we fix $(\beta, R) = (100, 0.1)$ and consider 6 different values of (k, m). For each choice of (k, m), Fig. 4.2 shows the empirical mean and a 95% confidence region when the curve $\ln(c_{(\beta,1,R)}/c_{(\beta,\gamma,R)})$, $0.1 \leq \gamma \leq 1$, has been estimated 100 times (independently of each other) by path sampling as described above. The computer time relative to the fastest case (the upper left plot where $(k, m) = (16, 0)$) is shown in each plot. The plots indicate that it suffices to have a small value of k but a large value of m. Reasonable good and fast estimates are obtained when $(k, m) = (16, 1000)$ and $(k, m) = (16, 10000)$.

Example 11: Fig. 4.3 shows a realisation x of a Strauss process on the unit square when $(\beta, \gamma, R) = (100, 0.5, 0.05)$ and produced using perfect simulation. We use this as our data in the following. Note that both $n(x) = 68$ and $s_{0.05}(x) = 3$ are rather small.

Here a rectangular 3D-grid of $76 \times 17 \times 10 = 12920$ points for (β, γ, R) in $[35, 110] \times [0.1, 1] \times [0.01, 0.1]$ has been used. The estimated likelihood function is rather flat with respect to β (not shown here). We obtain an approximate

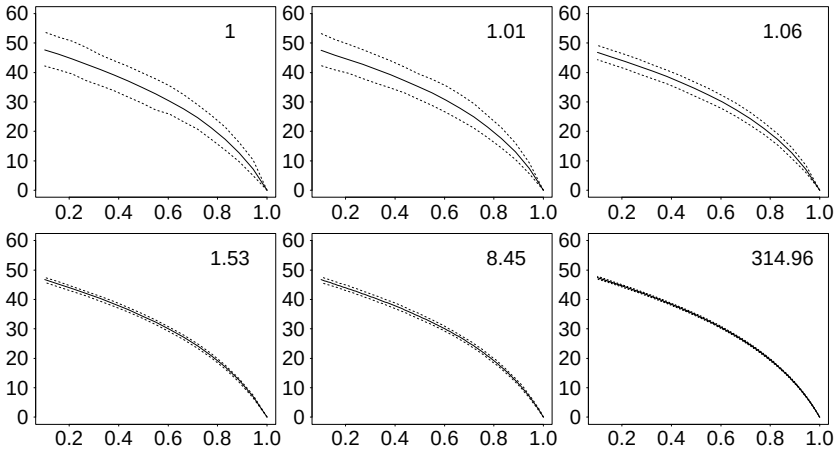

Figure 4.2: Empirical means of estimated curves $\gamma \to \ln(c_{(\beta,1,R)}/c_{(\beta,\gamma,R)})$ (solid lines) and 97.5% and 2.5% quantiles (dotted lines) when $(\beta, R) = (100, 0.1)$. The figures show the computer times relative to the fastest case. First row, from left to right: $(k, m) = (16, 0),\ (16, 100),\ (16, 1000)$. Second row, from left to right: $(k, m) = (16, 10000),\ (128, 1000),\ (4096, 0)$.

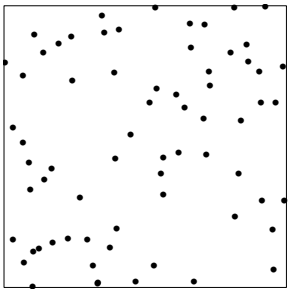

Figure 4.3: A realisation of the Strauss process on the unit square when $(\beta, \gamma, R) = (100, 0.5, 0.05)$ and obtained by perfect simulation.

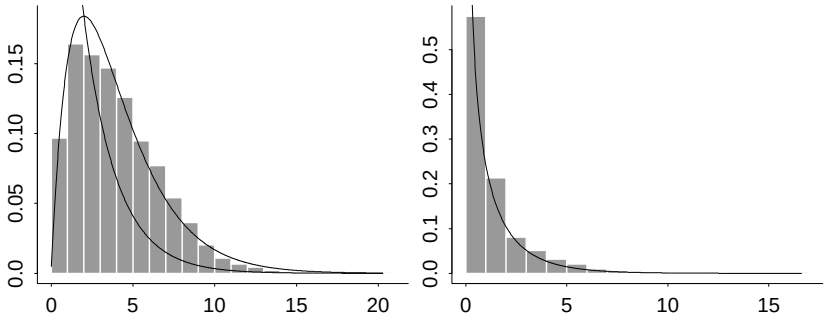

Figure 4.4: The empirical distribution of $-2\ln Q$ under the Poisson hypothesis $\gamma = 1$. Left plot: when all three parameters in the Strauss process are unknown; the densities of $\chi^2(2)$ and $\chi^2(4)$ are also shown. Right plot: fixed $R = 0.05$, and the density of $\chi^2(1)$.

MLE $(\hat{\beta}, \hat{\gamma}, \hat{R}) = (103, 0.16, 0.05)$, which except for $\hat{\gamma}$ is close to the true value; the low value of $\hat{\gamma}$ is caused by the low value of $s_{0.05}(x)$.

We have also investigated the empirical distribution for $(\hat{\beta}, \hat{\gamma}, \hat{R})$ obtained from 1000 independent perfect simulations with $(\beta, \gamma, R) = (100, 0.5, 0.05)$ (not shown here). The empirical marginal distributions have distinct modes about the true parameter values, and empirical means 100.88, 0.50, 0.057, respectively. The correlation is most pronounced between $\hat{\gamma}$ and $\hat{R}$: the empirical correlations are 0.17 for $(\hat{\beta}, \hat{\gamma})$, 0.24 for $(\hat{\beta}, \hat{R})$, and 0.80 for $(\hat{\gamma}, \hat{R})$.

Asymptotic normality of the MLE for the Strauss process with R fixed has been established in Jensen & Künsch (1994), but apart from that little seems to be known, including how the likelihood ratio test statistic for hypothesis testing is distributed. We now demonstrate that one should be careful with the use of "standard asymptotic results" and use simulations instead.

The left plot in Fig. 4.4 shows the empirical distribution of the log likelihood ratio statistic $-2\ln Q$ for the Poisson hypothesis $\gamma = 1$. It is obtained from 10000 independent simulations under the Poisson model with $\beta = 68$ (the MLE when $\gamma = 1$). The degrees of freedom between the Strauss and the Poisson model is $3 - 1 = 2$. As seen in the figure, $-2\ln Q$ is better described by a $\chi^2(4)$-distribution than a $\chi^2(2)$-distribution.

The Strauss model is only an exponential family model if we fix R; then the degrees of freedom between the Strauss and the Poisson model is $2 - 1 = 1$. For fixed $R = 0.05$, approximating the distribution of $-2\ln Q$ by a $\chi^2(1)$-distribution seems rather satisfactory, cf. the right plot in Fig. 4.4.

4.4 Non-parametric Bayesian MCMC inference for pairwise interaction point processes

In this section we discuss non-parametric Bayesian MCMC inference when the likelihood term is given by the pairwise interaction point process (4.1). In Section 4.4.1 we specify a prior for (β, φ), derive the posterior given an observation x, and point out the similarities and differences with the approach used in Heikkinen & Penttinen (1999). In Section 4.4.2 we propose an MCMC algorithm for the posterior. As ratios of normalising constants $c_{\beta,\varphi}$ have to be estimated in each update, path sampling is used during the simulations. The details are explained in Section 4.4.3, where also the use of perfect simulation is recalled. Finally, in Section 4.4.3 we consider some simulated and real datasets, and discuss the results obtained by either perfect or non-perfect MCMC simulation when the normalising constants are estimated.

4.4.1 Specification of prior and posterior

Assuming that φ is non-decreasing, we approximate φ by the step function (4.3). This assumption is in line with the way Arjas & Gasbarra (1994) specify a prior for non-parametric Bayesian estimation, observing in Section 4.4.3 that the posterior mean estimate of φ is a smooth curve. Rather than going into a detailed discussion on their idea and reasoning, let us just mention that our step function is chosen for pragmatic reasons: connected line segments or another type of splines could be used, but then the implementation of the posterior simulation procedure (Section 4.4.2) would be more complicated and slow down the posterior simulations. See also the discussion in Heikkinen & Penttinen (1999).

We use the following notation. The step function is given by the marked point process $\psi = \{(r_1, \gamma_1), \ldots, (r_p, \gamma_p)\}$, where the change-points are ordered $0 < r_1 < \ldots < r_p$ and their associated marks are ordered, assuming $0 < \gamma_1 < \ldots < \gamma_p < 1$. Note that r_p is the range of interaction. We let the number of change points $p \geq 0$ be random. Set $r_0 = 0$, $\gamma_0 = 0$, $\gamma_{p+1} = 1$, and $\delta_i = (\gamma_i - \gamma_{i-1})$, $i = 1, \ldots, p+1$. For p fixed, the state space of $(\delta_1, \ldots, \delta_{p+1})$ is the simplex in $\mathbb{R}^{p+1}$. It is convenient to transform $(\delta_1, \ldots, \delta_{p+1})$ into $(\zeta_1, \ldots, \zeta_p) = (\ln(\delta_2/\delta_1), \ldots, \ln(\delta_{p+1}/\delta_p))$ with state space $\mathbb{R}^p$.

We assume a priori that ψ and β are independent, and

(a) $r_1 < \ldots < r_p$ are the events of a homogeneous Poisson process on $[0, r_{\max}]$ of rate κ;

(b) conditionally on $(r_1, \ldots, r_p)$, we have that $\zeta_p, \ldots, \zeta_1$ is a Markov chain with $\zeta_p \sim \mathrm{N}(0, \sigma_p^2)$ and $\zeta_i | \zeta_{i+1} \sim \mathrm{N}(\zeta_{i+1}, \sigma_i^2)$, $i = p-1, \ldots, 1$;

(c) $\beta \sim \mathrm{Uniform}[\beta_{\min}, \beta_{\max}]$.

While (a) and (c) seem quite natural, (b) looks perhaps a bit arbitrary, but it is well-supported by the discussion below on (b1)–(b2).

In contrast to our model, Heikkinen & Penttinen (1999) fix p and $r_1, \ldots, r_p$ in the prior, and condition on $n(x) = n$ in the likelihood where n is the observed number of points. They assume that $\gamma_p, \ldots, \gamma_1$ is a Markov chain with $\ln \gamma_i | \ln \gamma_{i+1} \sim N(\ln \gamma_{i+1}, \sigma_i^2)$, $i = p, \ldots, 1$, where $\gamma_{p+1} = 1$ and the σ_i^2 are determined so that

$$\mathbb{V}\mathrm{ar}(\gamma_1|\gamma_2) = \ldots = \mathbb{V}\mathrm{ar}(\gamma_p|\gamma_{p+1}). \tag{4.12}$$

As noticed in Heikkinen & Penttinen (1999), some constraints such as $0 < \gamma_i \leq 1$, $i = 1, \ldots, p$, are needed in their prior model in order to obtain a well defined likelihood in the unconditional case.

The hyperparameters $r_{\max} > 0$, $\kappa > 0$, etc. in (a)–(c) should be chosen in accordance to background knowledge on (β, φ). The choice of $(r_{\max}, \kappa)$ specifies the "resolution", since $\mathbb{E}p = \kappa r_{\max}$. Often a crude estimate $\hat{R}$ of the range of interaction can be obtained (van Lieshout & Baddeley 1996). Then $r_{\max}$ should be chosen to be substantially larger than $\hat{R}$, so that $\varphi(r)$ is expected to be rather stable and close to 1 for large values of $r \leq r_{\max}$ (hence it is natural in (b) to generate the Markov chain backwards rather than forwards). The choice of variances in (b) may depend on the change-points, but in the following $\sigma_p^2 = \sigma^2$ is chosen not to depend on these, and for $i < p$, we allow σ_i^2 to depend only on ζ_{i+1} and (i, p); two specific models are studied in (b1) and (b2) below. Finally, in order to determine the hyperparameters $0 \leq \beta_{\min} < \beta_{\max}$, the expression $n(x)/|S|$ may be used as a lower bound on the MLE of β, where $|S|$ is the area of S.

The means for the Markov chain in (b) are specified so that $(\delta_2/\delta_1, \ldots, \delta_{p+1}/\delta_p)$ and $(\delta_1/\delta_2, \ldots, \delta_p/\delta_{p+1})$ are identically distributed. We have studied two models for the variances in detail:

(b1) $\sigma_1^2 = \ldots = \sigma_p^2 = \sigma^2$.

(b2) $\sigma_p^2 = \sigma^2$ and $\sigma_i^2 = \ln[1/2 + \{(\sigma\delta_{i+1}/\delta_{i+2})^2 + 1/4\}^{1/2}]$, $i = 1, \ldots, p-1$.

Here $\sigma > 0$ is a tuning parameter. Under (b1) the conditional distribution of $(\zeta_1, \ldots, \zeta_p)$ given $(r_1, \ldots, r_p)$ is a multivariate normal distribution with mean zero, variances $\mathbb{V}\mathrm{ar}(\zeta_i) = (p-i+1)\sigma^2$, $1 \leq i \leq p$, and correlations $\mathbb{C}\mathrm{orr}(\zeta_{i-j}, \zeta_i) = ((p-i+1)/(p-i+j+1))^{1/2}$, $1 \leq j < i \leq p$. Thus $(\delta_1, \ldots, \delta_{p+1})$ follows an additive logistic normal distribution (Aitchison 1986). The conditional distribution of $(\zeta_1, \ldots, \zeta_p)$ is more complicated under (b2). Here the σ_i^2 are determined so that

$$\mathbb{V}\mathrm{ar}(\delta_2/\delta_1|\delta_3/\delta_2) = \ldots = \mathbb{V}\mathrm{ar}(\delta_p/\delta_{p-1}|\delta_{p+1}/\delta_p) = \mathbb{V}\mathrm{ar}(\delta_{p+1}/\delta_p),$$

compare with (4.12).

Simulated results for the prior distribution of φ look rather similar when different values of σ are used in the two models (b1) and (b2), so the choice of σ is important while it is less important which model is used. Fig. 4.5 shows the smoothing effect of σ in (b1) when $\kappa r_{\max} = 5$.

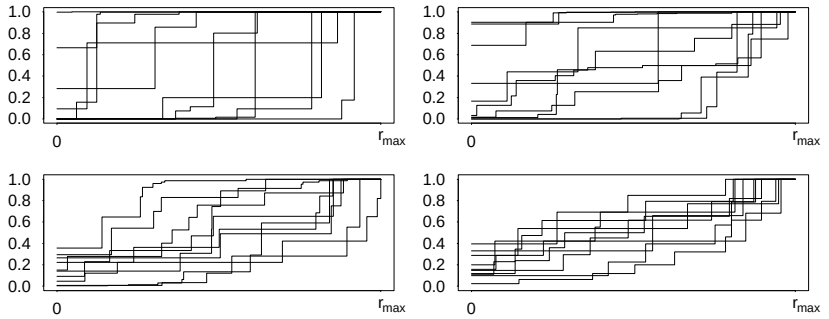

Figure 4.5: Each plot shows ten independent realizations of φ under (b1). From top left to bottom right: $\sigma^2 = 10, 1, 0.1, 0.01$.

Finally, we derive the posterior density. The Jacobian of the transformation $(\zeta_1, \ldots, \zeta_p) \to (\gamma_1, \ldots, \gamma_p)$ is

$$|\partial(\gamma_1, \ldots, \gamma_p)/\partial(\zeta_1, \ldots, \zeta_p)| = \delta_1 \times \cdots \times \delta_{p+1}.$$

Let μ denote the unit rate Poisson process on $[0, r_{\max}] \times [0, 1]$. Then under (a)–(b), $\psi = \{(r_1, \gamma_1), \ldots, (r_p, \gamma_p)\}$ with $r_1 < \ldots < r_p$ has prior density

$$\begin{aligned} \pi(\psi) \quad &\propto \quad \kappa^p \mathbf{1}[0 < \gamma_1 < \ldots < \gamma_p < 1]/(\delta_1 \times \cdots \times \delta_{p+1}) \\ &\times \quad \prod_{i=1}^{p} \left(2\pi\sigma_i^2\right)^{-1/2} \exp\left(- \left(\zeta_i - \zeta_{i+1}\right)^2 / \left(2\sigma_i^2\right)\right) \end{aligned} \tag{4.13}$$

with respect to μ, where we set $\zeta_{p+1} = 0$. Setting

$$s_i = \sum_{\{\xi,\eta\} \subseteq x} \mathbf{1}[r_{i-1} < \|\xi - \eta\| \leq r_i], \quad i = 1, \ldots, p,$$

the posterior density of $\theta = (\beta, \psi)$ is

$$\pi(\theta|x) \propto \pi(\psi)\beta^{n(x)}\gamma_1^{s_1} \times \cdots \times \gamma_p^{s_p}/c_\theta \tag{4.14}$$

with respect to the product measure of the Lebesgue measure on $[\beta_{\min}, \beta_{\max}]$ and μ, where $c_\theta = c_{\beta,\varphi}$ denotes the normalising constant of the likelihood.

4.4.2 Simulation of the posterior

The grid type approach in Section 3 is sensitive to the curse of dimensionality, and should preferably be avoided in the posterior simulations where the dimension of φ can increase considerably. We propose therefore to use a hybrid Metropolis-Hastings (also called Metropolis-within-Gibbs) algorithm for the posterior (4.14), where we alternate between updating β and ψ, and use in

Section 4.4.3 path sampling for estimating ratios of unknown normalising constants. In the description below of both types of updates we let (β, ψ) denote the current state of the algorithm.

For a β update, we use a Metropolis random walk algorithm: generate $\beta' \sim \text{Uniform}\big([\max\{\beta_{\min}, \beta - \epsilon\}, \min\{\beta_{\max}, \beta + \epsilon\}]\big)$, and replace β by β' with probability $\min\{1, \alpha_1(\beta, \beta', \psi)\}$, where

$$\alpha_1(\beta, \beta', \psi) = \frac{c_{\beta,\psi}}{c_{\beta',\psi}} \left(\frac{\beta'}{\beta}\right)^{n(x)} \frac{\min\{\beta_{\max}, \beta + \epsilon\} - \max\{\beta_{\min}, \beta - \epsilon\}}{\min\{\beta_{\max}, \beta' + \epsilon\} - \max\{\beta_{\min}, \beta' - \epsilon\}} \tag{4.15}$$

where $\epsilon > 0$ is a tuning parameter; retain β otherwise.

For a ψ update, we use a reversible jump MCMC algorithm (Green 1995) which essentially is a Metropolis-Hastings algorithm of the type studied in Geyer & Møller (1994): if $\psi = \{(r_1, \gamma_1), \ldots, (r_p, \gamma_p)\}$, then

- with probability 1/2 we make a birth proposal: generate $r' \sim \text{Uniform}[0, r_{\max}]$; find the interval where $r_{i-1} < r' \le r_i$ (setting $r_{p+1} = r_{\max}$); generate $\gamma' \sim \text{Uniform}[\gamma_{i-1}, \gamma_i]$; replace ψ by $\psi' = \psi \cup \{(r', \gamma')\}$ with probability $\min\{1, \alpha_2(\beta, \psi, \psi')\}$, where

$$\alpha_2(\beta, \psi, \psi') = \frac{c_{\beta,\psi}}{c_{\beta,\psi'}} \left(\frac{\gamma'}{\gamma_i}\right)^{s'} \frac{\pi(\psi')}{\pi(\psi)} \frac{r_{\max}(\gamma_i - \gamma_{i-1})}{p+1} \tag{4.16}$$

  with

$$s' = \sum_{\{\xi,\eta\} \subseteq x} \mathbf{1}[r_{i-1} < \|\xi - \eta\| \le r'];$$

  retain ψ otherwise;

- else we generate a death proposal: generate randomly uniformly $(r_i, \gamma_i) \in \psi$ (if $\psi = \emptyset$ we set $\{r_i, \gamma_i)\} = \emptyset$); replace ψ by $\psi' = \psi \setminus \{(r_i, \gamma_i)\}$ with probability $\min\{1, \alpha_2(\beta, \psi', \psi)^{-1}\}$; retain ψ otherwise.

Under (b1), $\pi(\psi')/\pi(\psi)$ in (4.16) depends only on $\gamma_{i-2}, \gamma_{i-1}, \gamma', \gamma_i, \gamma_{i+1}$ (with obvious adjustments at the boundaries $i = 1$ and $i = p$), cf. (4.13). For a β-update, if for example $\beta < \beta'$, the term

$$\ln(c_{\beta',\psi}/c_{\beta,\psi}) = \int_{\beta}^{\beta'} \mathbb{E}_{\tilde{\beta},\psi} n/\tilde{\beta} \, \mathrm{d}\tilde{\beta} \tag{4.17}$$

in (4.15) has to be estimated; and for a ψ-update, the term

$$\ln(c_{\beta,\psi\cup(r',\gamma')}/c_{\beta,\psi}) = \int_{\gamma_{i-1}}^{\gamma'} \mathbb{E}_{\beta,\psi\cup(r',\gamma')}(s_i - s')/\gamma \, \mathrm{d}\gamma \tag{4.18}$$

in (4.16) has to be estimated. We can do this by path sampling as discussed below.

4.4.3 Empirical results

We illustrate now how our method applies for two datasets when the prior for (φ, β) is specified by (a)–(c) and (b1).

Example 13 concerns the locations of 152 displaced amacrine cells within a rectangular $1070 \times 600\,\mu m^2$ region, see Fig. 4.6. This dataset has been analysed in many papers, including Diggle & Gratton (1984) and Heikkinen & Penttinen (1999). Diggle & Gratton (1984) modelled the data by the interaction function

$$\varphi_{\mathrm{DG}}(r) = \begin{cases} 0, & r < \sigma_{\mathrm{DG}} \\ ((r - \sigma_{\mathrm{DG}})/(\rho_{\mathrm{DG}} - \sigma_{\mathrm{DG}}))^{\beta_{\mathrm{DG}}}, & \sigma_{\mathrm{DG}} \le r \le \rho_{\mathrm{DG}} \\ 1, & r > \rho_{\mathrm{DG}}. \end{cases} \tag{4.19}$$

Note that σ_{DG} is a hard core parameter and ρ_{DG} specifies the range of interaction. Diggle & Gratton (1984) obtained the estimate

$$(\sigma_{\mathrm{DG}}, \rho_{\mathrm{DG}}, \beta_{\mathrm{DG}}) = (19, 76, 1.67)$$

by an ad hoc method.

When using path sampling we have to estimate the means of n and $s_i - s'$ from (4.17) and (4.18). We have tried to do this along similar lines as in Section 4.3.1. To examine whether perfect simulation is feasible under the estimated Diggle-Gratton model, we have considered a plot of $\mathbb{E}T$ versus β (not shown here). It is similar to the left plot in Fig. 4.1, but with a critical value for β around $200/(1070 \times 600)$. Simulated point patterns for $\beta = 200/(1070 \times 600)$, $(\sigma_{\mathrm{DG}}, \rho_{\mathrm{DG}}, \beta_{\mathrm{DG}}) = (19, 76, 1.67)$, and $S = [0, 1070] \times [0, 600]$ contain on average about 65 points. This is far below the 152 points in the data, so perfect simulation is not feasible. Instead we let the initial state be given by the result after an appropriate burn-in (again the sample is generated by the Metropolis-Hastings algorithm in Geyer & Møller (1994) but with a multiscale process as equilibrium distribution).

To check this procedure in a simplified situation, we discuss in Example 12 the results of the posterior analysis when using either a perfect simulation or an appropriate burn-in. The dataset in Example 12 is a perfect simulation of the Diggle-Gratton model but with $\beta = 100$, $(\sigma_{\mathrm{DG}}, \rho_{\mathrm{DG}}, \beta_{\mathrm{DG}}) = (0.025, 0.1, 1.67)$, and $S = [0, 1]^2$, see Fig. 4.6. This point pattern contains only 48 points, and compared to the amacrine cell data it is less regular.

Example 12: For the simulated data we let $(k, m) = (10, 1000)$. Compared to the values of k used in Fig. 4.2, it suffices to use the present smaller value of k, since we usually integrate over a shorter interval in (4.18). Furthermore, we choose $r_{\max} = 0.14$, $\kappa r_{\max} = 5$, $\sigma^2 = 1$, $\beta_{\min} = 50$, $\beta_{\max} = 115$, and $\epsilon = 2.5$. Time series plots for different statistics (not shown here) show that the algorithm for posterior simulations is mixing well, with an appropriate burn-in of only a few thousand updates, using an initial state of ψ where the r_i are generated from the Poisson prior and the δ_i are all equal.

The left plot in Fig. 4.7 shows some results based on a sample of 150000 posterior realizations of (ψ, β).

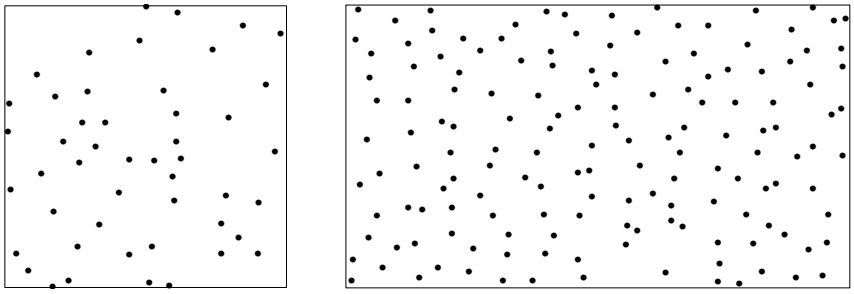

Figure 4.6: Left: simulated data. Right: amacrine cell data.

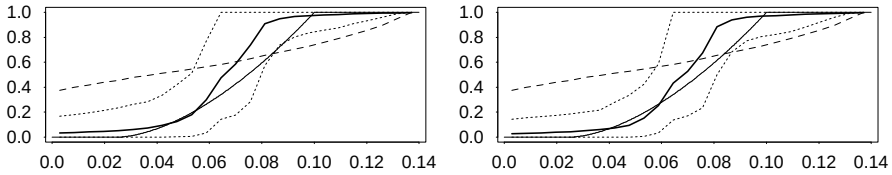

Figure 4.7: Results for the simulated data. Both plots: posterior mean (thick solid line) together with a 95% credibility interval (dotted lines), prior mean (dashed line), and true interaction function (thin solid line). Left: when perfect simulation and path sampling for estimating ratios of normalising constants is used. Right: when an appropriate burn-in is used instead.

There is a clear difference between the posterior and prior mean of $\varphi(r)$, $0 < r < r_{\max}$. The posterior mean of $\varphi(r)$ is rather close to $\varphi_{DG}(r)$ except for intermediate values of r where it clearly overestimates the true interaction function. However, the credibility interval for $\varphi(r)$ is rather wide, especially for intermediate values of r, and when we repeated everything for another simulated dataset (not shown here) the posterior mean did not overestimate the true φ.

Plots (not shown here) of the prior and empirical posterior distributions of r_1, r_p and p show the following. The posterior density of r_1 is close to the prior density $\pi(r_1) \propto \exp(-\kappa r_1)$ $(0 < r_1 < r_{\max})$ at least for small and modest values of r_1. The posterior and prior mean of r_1 are 0.026, and 0.027, which are both close to the hard core $\sigma_{DG} = 0.025$. There is a pronounced difference between the posterior density of r_p and the prior density $\pi(r_p) \propto \exp(-\kappa(r_{\max} - r_p))$ $(0 < r_p < r_{\max})$. The posterior and prior mean of r_p are 0.087 and 0.113, i.e. they are equally far from the true interaction range $\rho_{DG} = 0.1$. The prior and the posterior distribution for p are rather close. The posterior and prior mean are 5.12 and 5.

As mentioned we have for comparison also used an appropriate burn-in when

estimating ratios of normalising constants. With a burn-in of 250 iterations the results shown in the right plot of Fig. 4.7 are very close to those in the left plot in Fig. 4.7. For comparison, the average coalescence time of the perfect simulation algorithm (with parameters as used for the simulated data in Fig. 4.6) is 7.73×10^3. In conclusion we find it both safe and much faster to use the "non-perfect" simulations.

Example 13: For the amacrine cell data we let $(k, m) = (10, 2500)$, $r_{\max} = 120$, $\kappa r_{\max} = 5$, $\sigma^2 = 1$, $\beta_{\min} = 0.005$, $\beta_{\max} = 0.05$, and $\epsilon = 0.0005$. As mentioned we have to use non-perfect simulation when estimating ratios of normalising constants; for this we use a burn-in of 2500 iterations. For instance, when simulating a multiscale process which approximates the estimated Diggle-Gratton model, the Metropolis-Hastings chain seems to be stabilised after 1500 iterations when started in a realization from a homogeneous Poisson process on $[0, 1070] \times [0, 600]$ with mean number of points equal to 152. For the algorithm for posterior simulations, time series plots (not shown here) for different statistics show that the algorithm converges quickly into equilibrium (using a similar initial state as in Example 12). The mixing properties of this algorithm are sensitive to the choice of σ; it is slowly mixing for $\sigma = 0.1$, while it is faster mixing for $\sigma = 1$ and $\sigma = 10$. The posterior results for $\sigma = 10$ are very close to the corresponding results for $\sigma = 1$. In the sequel we report on results obtained with $\sigma = 1$.

Fig. 4.8 shows some results for the amacrine cell data based on 250000 posterior updates. The top left plot in Fig. 4.8 is similar to the plots in Fig. 4.7. Since we do not know the truth, we have shown the estimated Diggle-Gratton interaction function instead. Again there is a clear difference between the posterior and prior mean of $\varphi(r)$. The plot may be compared with Fig. 3 in Heikkinen & Penttinen (1999) which shows their maximum a posteriori estimate of φ and several other estimated interaction functions from the literature. Most but not all of these estimated interaction functions are within the 95% credibility interval in Fig. 4.8. The maximum a posteriori estimate is close to $\varphi_{\mathrm{DG}}(r)$ when $r \leq 76$, and larger than 1 for $r > 76$. Furthermore, the Heikkinen-Penttinen maximum a posteriori estimate and our posterior estimate exhibit a similar behaviour for $r \leq 60$. The "steps" in our estimate may be a consequence of the fact that the amacrine cell data are regular; such steps are not so pronounced in the Heikkinen-Penttinen estimate, possibly due to the choice of smoothing parameter and because they connect the midpoints of the estimated steps with straight line segments.

The top right plot in Fig. 4.8 shows that the posterior density of r_1 is rather flat for $r_1 \leq 20$ and has a mode for $r_1 > 20$, while the prior for r_1 is strictly decreasing.

The lower left plot in Fig. 4.8 shows a large dispersion in the posterior distribution for r_p, indicating difficulties in estimating the range of interaction; the value $\rho_{\mathrm{DG}} = 76$ obtained by Diggle & Gratton (1984) is in the centre of the posterior distribution.

The lower right plot in Fig. 4.8 shows clearly that the posterior and prior

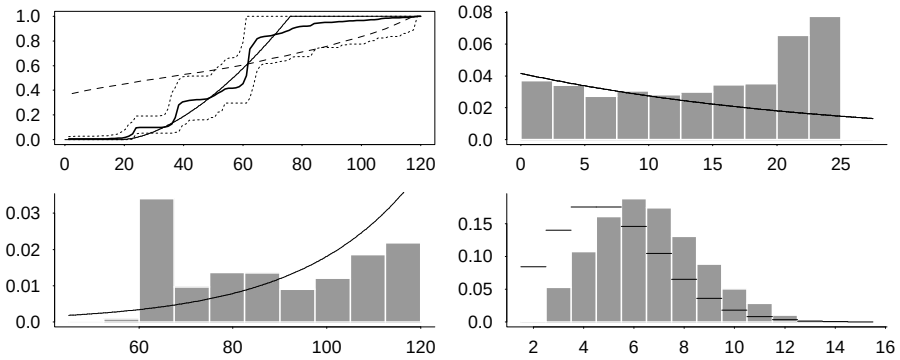

Figure 4.8: Results for the amacrine data. Top left: as Fig. 4.7. Top right, bottom left, and bottom right: posterior and prior density (full line) of r_1, r_p, and p.

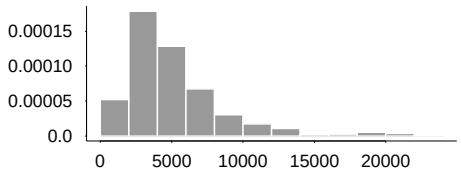

Figure 4.9: Posterior distribution of $\beta \times 1070 \times 600$.

distribution for p are different. Compared to Example 12, we need now a larger number of change-points due to the increased amount of data, but the number needed is still small compared to the 30 change-points used in Fig. 3 in Heikkinen & Penttinen (1999). The posterior and prior means of p are 6.62 and 5.

Finally, Fig. 4.9 shows the posterior distribution of $\beta \times 1070 \times 600$, with a mode around 4000, a mean of about 4589, and a large dispersion. In order to obtain about 152 points in average, simulation of the estimated Diggle-Gratton model requires $\beta \times 1070 \times 600 \approx 6500$, which is larger than the posterior mean but still not in the tail of the posterior distribution.

4.5 Concluding remarks

In this paper we have considered the combination of path sampling and perfect simulation in the context of both likelihood inference (Section 4.3) and non-parametric inference (Section 4.4) for pairwise interaction point processes. We now summarise the merits and limitations of our approach.

The feasibility of perfect simulation was illustrated in Example 9 and Sec-

tion 4.4.3. Particularly, perfect simulation is (so far) not feasible for the amacrine cell data, cf. the beginning and end of Section 4.4.3. However, we have so far not exploited the possibility of speeding things up by a parallel implementation of the path sampling procedure (running k chains in parallel). Access to increasing amount of computing power will of course also be in favour of perfect simulation.

When perfect simulation is feasible, we have proposed to use this for the initial state of each Markov chain in connection to the path sampling procedure (Sections 4.3.1 and 4.4.3). We find this attractive for several reasons: If the initial state is not in equilibrium, we need to determine an appropriate burn-in before we start sampling. As the rate of convergence for the Metropolis-Hastings algorithm depends much on the value of θ, so should the burn-in. At least to our knowledge there are no analytical results (Møller 1999) or automatic methods (apart from perfect simulation) for determining an appropriate burn-in. Particularly, for the likelihood surface considered in Section 4.3, we have to perform a thorough output analysis for each value of θ in the 3D-grid, and this can be rather time-consuming. Moreover, in general we cannot be assured if the burn-in is appropriately determined. However, when X_0 is generated by the dominated CFTP algorithm, we let the machine do all the work so that we are certain that the sample $X_i \sim f_\theta,\ i = 0, \ldots, m$, is in equilibrium. Finally, perfect simulation allows us to check a method based on non-perfect simulation; we illustrated this point in Example 12.

If perfect simulation is feasible, one may expect that the Metropolis-Hastings algorithm has good mixing properties (though the birth-and-death process underlying the perfect sampler in Section 4.2 is different from that used for the Metropolis-Hastings algorithm). Our empirical evidence support this. Both the Metropolis-Hastings algorithm and that based on birth-and-death processes converge geometrically fast (Møller 1989, Geyer & Møller 1994, Geyer 1999), but still it remains to obtain a deeper theoretical understanding of the mixing properties, cf. Møller (1999).

In this paper we have used the perhaps simplest version of path sampling. Another is based on simulated tempering (Marinari & Parisi 1992, Geyer & Thompson 1995, Gelman & Meng 1998). The advantage of this could be that only one burn-in or perfect simulation is needed each time we do path sampling. On the other hand, it suffices at least for the examples we have considered to use only $k = 10$–16 burn-in or perfect simulations and the simulated tempering procedure will involve some extra work when determining an appropriate "pseudo-prior". Finally, we would expect that perfect simulation for simulated tempering as described in Møller & Nicholls (1999) and Brooks et al. (2002) may only apply in cases where one of the models in the paths in (4.17) and (4.18) are sufficiently close to a Poisson process.

Acknowledgement

This research will be a part of KKB's PhD dissertation. JM was supported by the European Union's research network "Statistical and Computational Methods for the Analysis of Spatial Data, ERB-FMRX-CT96-0096", by the Centre for Mathematical Physics and Stochastics (MaPhySto), funded by a grant from the Danish National Research Foundation, and by the Danish Natural Science Research Council. We are grateful for the constructive comments made by an associate editor and two referees.

Bibliography

Aitchison, J. (1986). *The statistical analysis of compositional data*. Monographs on Statistics and Applied Probability, Chapman and Hall, London.

Arjas, E. & Gasbarra, D. (1994). Nonparametric Bayesian inference from right censored survival data, using the Gibbs sampler. *Statist. Sinica* **4**, 505–524.

Baddeley, A. & Turner, R. (2000). Practical maximum pseudolikelihood for spatial point patterns. *Australian and New Zealand Journal of Statistics* **42**, 283–322.

Berthelsen, K. K. & Møller, J. (2002). A primer on perfect simulation for spatial point processes. *Bulletin of the Brazilian Mathematical Society* **33**. To appear.

Brooks, S. P., Fan, Y. & Rosenthal, J. S. (2002). Perfect forward simulation via simulated tempering. Submitted.

Diggle, P. & Gratton, R. (1984). Monte Carlo methods of inference for implicit statistical models (with discussion). *J. Roy. Statist. Soc. Ser. B* **46**, 193–227.

Gates, D. J. & Westcott, M. (1986), Clustering estimates for spatial point distributions with unstable potentials. *Ann. Inst. Statist. Math.* **A 38**, 123–135.

Gelman, A. & Meng, X.-L. (1998). Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. *Statist. Sci.* **13**, 163–185.

Geyer, C. J. (1999). Likelihood inference for spatial processes. In *Stochastic geometry, likelihood and computation* (eds O. E. Barndorff-Nielsen, W. Kendall & M. N. M. van Lieshout), pp. 79–140. Chapman & Hall.

Geyer, C. J. & Møller, J. (1994), Simulation procedures and likelihood inference for spatial point processes. *Scand. J. Statist.* **21**, 359–373.

Geyer, C. J. & Thompson, E. A. (1995). Annealing Markov chain Monte Carlo with applications to ancestral inference. *J. Amer. Statist. Assoc.* **90**, 909–920.

Green, P. (1995). Reversible jump MCMC computation and Bayesian model determination. *Biometrika* **82**, 711–732.

Heikkinen, J. & Penttinen, A. (1999). Bayesian smoothing in the eastimation of the pair potential function of Gibbs point processes. *Bernoulli* **5**, 1119–1136.

Jensen, J. L. & Künsch, H. (1994). On asymptotic normality of pseudo likelihood estimates for pairwise interaction processes. *Ann. Inst. Statist. Math.* **46**, 475–486.

Kelly, F. P. & Ripley, B. D. (1976). A note on Strauss's model for clustering. *Biometrika* **63**, 357–360.

Kendall, W. S. (1998). Perfect simulation for the area-interaction point process. In *Probability towards 2000* (eds L. Accardi & C. Heyde), pp. 218–234. Springer, New York.

Kendall, W. S. & Møller, J. (2000). Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes. *Adv. Appl. Prob.* **32**, 844–865.

Lieshout, M. N. M. van & Baddeley, A. J. (1996). A nonparametric measure of spatial interaction in point patterns. *Statist. Neerlandica* **50**, 344–361.

Lieshout, M. N. M. van & Zwet, E. W. van (2001). Exact sampling from conditional bolean models with applications to maximum likelihood inference. *Adv. Appl. Prob.* **33**, 339–353.

Loizeaux, M. A. & McKeague, I. W. (2001). Perfect sampling for posterior landmark distributions with application to the detection of disease clusters. In *Selected proceedings of the symposium on inference for stochastic processes* (eds I. V. Basawa, C. C. Heyde & R. L. Taylor), Vol. 37, IMS Lecture Notes & Monographs Series, pp. 321–331.

Lund, J. & Thönnes, E. (2000). Perfect simulation for point processes given noisy observations. Research Report 366, Department of Statistics, University of Warwick.

Marinari, E. & Parisi, G. (1992). Simulated tempering: A new Monte Carlo scheme. *Europhysics letters* **19**, 451–458.

McKeague, I. W. & Loizeaux, M. A. (2002). Perfect sampling for point process cluster modelling. In *Spatial cluster modelling* (eds A. Lawson & D. Denison) pp. 87–107. Chapman & Hall/CRC, Boca Raton.

Møller, J. (1989). On the rate of convergence of spatial birth-and-death processes. *Ann. Inst. Statist. Math.* **41**, 565–581.

Møller, J. (1999). Markov chain Monte Carlo and spatial point processes. In *Stochastic geometry, likelihood and computation* (eds O. E. Barndorff-Nielsen, W. S. Kendall & M. N. M. van Lieshout), pp. 141–172. Chapman & Hall.

Møller, J. (2001). A review of perfect simulation in stochastic geometry. In *Selected proceedings of the symposium on inference for stochastic processes* (eds I. V. Basawa, C. C. Heyde & R. L. Taylor), Vol. 37, IMS Lecture Notes & Monographs Series, pp. 333–355.

Møller, J. & Nicholls, G. (1999). Perfect simulation for sample-based inference. Technical report, R-99-2011, Department of Mathematical Sciences, Aalborg University.

Penttinen, A. (1984). *Modelling interaction in spatial point patterns: Parameter estimation by the maximum likelihood method*. Number 7 in Jyväskylä Studies in Computer Science, Economics, and Statistics.

Propp, J. G. & Wilson, D. B. (1996), Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms* **9**, 223–252.

Ripley, B. D. (1988). *Statistical inference for spatial processes*. Cambridge University Press, Cambridge.

Ripley, B. D. (1989). Gibbsian interaction models. In *Spatial statistics: past, present and future* (ed D. A. Griffiths), pp. 1–19. Image, New York.

Strauss, D. J. (1975). A model for clustering. *Biometrika* **62**, 467–475.

5 Markov chain Monte Carlo without the normalising constant

Kasper K. Berthelsen and Jesper Møller
Aalborg University

Anthony N. Pettitt and Robert Reeves
Queensland University of Technology

Abstract. A recurring problem when using a Metropolis-Hastings algorithm for simulating the posterior distribution in a Bayesian setup is that it involves evaluating ratios of unknown normalising constants. We consider a method that avoids the ratio of unknown normalising constants by introducing an auxiliary variable. This auxiliary variable method requires that it is possible to sample from the likelihood. Further, the success of the method relies on how well the auxiliary variable approximates the likelihood. This is illustrated in an example in the setting of spatial point processes.

5.1 Introduction

In many applications of Metropolis-Hastings algorithms it is necessary to approximate ratios of unknown normalising constants. Traditional ways of approximating such ratios of unknown normalising constants include importance sampling, bridge sampling (Meng & Wong 1996) and path sampling (Gelman & Meng 1998). In this paper we consider a method which avoids such approximations. The price one has to pay is the introduction of an auxiliary process specified by a density with a known normalising constant. Further, this auxiliary variable method requires exact samples from the likelihood — at least for the range of parameters of interest. As we shall see the feasibility of this method relies on specifying an auxiliary process that approximates the likelihood sufficiently well.

Consider a Bayesian setup and assume that the likelihood is given by $\pi(x|\theta) = Z_\theta^{-1} q_\theta(x)$ where Z_θ is the unknown normalising constant. Further, assume that the normalised prior on θ is $\pi(\theta)$. Given data y we want to sample from the posterior $\pi(\theta|y) = \pi(y|\theta)\pi(\theta)$. Going about it the usual way using a Metropolis-Hastings algorithm this involves evaluating ratios of unknown normalising constants. Instead introduce an auxiliary process with normalised density $f(\cdot|\theta, y)$ which is defined on the same space as $\pi(\cdot|\theta)$. The idea is then to sample (θ, x)

from

$$f(x|\theta, y)\pi(\theta|y) \tag{5.1}$$

noting that the marginal distribution of θ has density $\pi(y|\theta)\pi(\theta)$.

Simulating from (5.1) is done using a Metropolis-Hastings algorithm. Assume that the proposal distribution can be written as

$$p(x', \theta'|x, \theta, y) = \pi(x'|\theta')p(\theta'|\theta, y)$$

where $p(\cdot|\theta, y)$ is a normalised density. Then we can generate a realisation from the proposal distribution by first generating θ' from $p(\cdot|\theta, y)$ and then generate x' from the likelihood $\pi(\cdot|\theta')$. The Hastings ratio is then

$$H(\theta', x'|\theta, x, y) = \frac{f(x'|\theta', y)\pi(y|\theta')\pi(\theta')\pi(x|\theta)p(\theta|\theta', y)}{f(x|\theta, y)\pi(y|\theta)\pi(\theta)\pi(x'|\theta')p(\theta'|\theta, y)}. \tag{5.2}$$

The key feature of (5.2) is that the unknown normalising constants cancel.

Note that if $f(x|\theta, y) = \pi(y|\theta)$ then we are back in the conventional case. Assume that in the conventional case the Metropolis-Hastings algorithm is mixing well (using some method for approximating the ratios of unknown normalising constants). Then the auxiliary variable method is expected to work well too if the auxiliary process is a sufficiently good approximation of the likelihood.

5.2 Auxiliary variable method for point processes

We now consider the auxiliary variable method in the setting of spatial point processes. We restrict attention to a Strauss point process (Strauss 1975, Kelly & Ripley 1976) defined on a bounded region $S \subset \mathbb{R}^2$ by a density

$$\pi(x|\beta, \gamma, R) = \frac{1}{Z_\theta}\beta^{n(x)}\gamma^{s_R(x)} \tag{5.3}$$

w.r.t. a unit rate Poisson point process on S. Here x is a finite subset of S referred to as a point configuration, and Z_θ is an unknown normalising constant, where $\theta = (\beta, \gamma, R)$. Further, $\beta > 0$ is the intensity (also known as the chemical activity in statistical physics), $n(x)$ is the cardinality of x, $0 < \gamma \leq 1$ is the interaction parameter, $R > 0$ is the interaction range, and

$$s_R(x) = \sum_{\{\xi,\eta\}\subseteq x:\xi\neq\eta} \mathbf{1}[\|\eta - \xi\| \leq R]$$

is the number of pairs of points in x within a distance R from each other. In the analysis below we have for simplicity fixed $R = 0.05$ but it is straightforward to extend everything to the case of variable interaction range. Realisations of a Strauss point process are in general regular compared to those of a Poisson point process due to the inhibition in the model. See Figure 5.1 for a realisation of a Strauss point process.

A Strauss point process is an example of a locally stable point process which can be simulated exactly by dominated coupling from the past (dominated CFTP). For more details on dominated CFTP in general and the application to locally stable point processes in particular, see Kendall & Møller (2000).

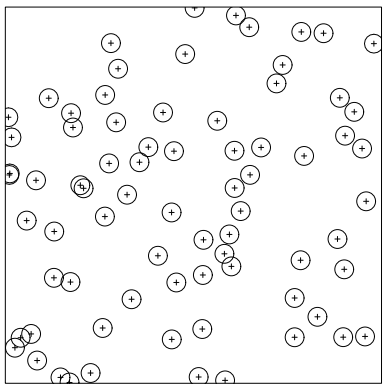

Figure 5.1: Realisation of a Strauss point process with parameter $(\beta, \gamma, R) = (100, 0.5, 0.05)$ on the unit square. Circles centred at points have radii 0.025.

5.2.1 Auxiliary point processes

In this section we consider three auxiliary point processes.

Let μ be a unit rate Poisson point process on S. The simplest choice of an auxiliary point process is a Poisson point process on S with intensity κ which has a normalised density $f(x|\kappa) = e^{(1-\kappa)|S|}\kappa^{n(x)}$ w.r.t. μ, where $|S|$ is the area of the region S. We need to specify how κ should depend on y and θ. We let $\kappa = n(y)/|S|$ be the MLE of κ for the data y, implying an auxiliary point process with density

$$f(x|\theta, y) = e^{|S|-n(y)}\left(\frac{n(y)}{|S|}\right)^{n(x)} \tag{5.4}$$

w.r.t. μ. We refer to this auxiliary process as the fixed Poisson process. In the simulation study in Section 5.2.2 we see that this choice of auxiliary process is far from good enough. This is not surprising as (5.4) does not take the pairwise interaction or the value of θ into account.

The second auxiliary process is chosen to have a fixed density $\pi(\cdot|\hat{\theta})$ where $\hat{\theta}$ is the MLE of θ for the data y. This approach implies an auxiliary process with density

$$f(x|\theta, y) \propto \hat{\beta}^{n(x)}\hat{\gamma}^{s_R(x)}, \tag{5.5}$$

w.r.t. μ where $(\hat{\beta}, \hat{\gamma}) = \operatorname{argmax}_{(\tilde{\beta},\tilde{\gamma})}\pi(y|\tilde{\beta}, \tilde{\gamma}, R)$ is the MLE. We refer to this process as the fixed Strauss process. Note that (5.5) only depends on y and not on θ. Further, because the density (5.5) is fixed we do not need to know the normalising constant. Due to available software we have chosen to find the MLE by approximating the likelihood function on a grid of β and γ values using path sampling (Gelman & Meng 1998). Alternatively one could have used pseudo

likelihood (Besag 1975, Ripley 1988) or MCMCMLE (Penttinen 1984, Geyer & Thompson 1992). It does not seem crucial that this MLE is very accurate as it only serves as the parameter in the auxiliary process. Not surprisingly this fixed Strauss process works much better than the fixed Poisson processes as it takes the pairwise interaction into account. Still it does not depend on the parameters of the Strauss process.

The third auxiliary point process we consider takes both pairwise interaction and parameters into account. This point process is an example of a so-called partially ordered Markov model (POMM). POMMs have the attractive properties that their normalising constants are known (and equal one), and that they can model some degree of interaction. The idea of a POMM type point process is best described by how to generate a realisation of it. Initially divide the region S into a finite number of disjoint subsets in the sequel referred to as cells. The simulation is then done in a single sweep where the configuration of each cell is only updated once in some (possibly random) order. The update of the configuration of a given cell only depends on the configuration of the cells already updated. The POMM class of lattice process models was introduced by Cressie & Davidson (1998) and Davidson, Cressie & Hua (1999) who applied POMMs in the analysis of grey scaled digital images. Cressie, Zhu, Baddeley & Nair (2000) consider directed Markov point processes (DMPP) as limits of POMMs. They use a Strauss like DMPP which suffers from clear directional effects that do not show up in the examples they consider. Our POMM resembles the POMM used in their paper but does not suffer from any apparent directional effects.

The POMM we consider is specified as follows. Let I be a finite index set with an ordering $\prec$ where $i \prec j$ implies $i \neq j$. Let $\{C_i \subseteq S : i \in I\}$ be a division of S into disjoint cells, i.e. $C_i \cap C_j = \emptyset$ for $i \neq j$ and $S = \cup_{i \in I} C_i$. Further, let $\rho : I \mapsto I$ be a random permutation of I with an associated ordering $\prec_\rho$ of I defined by $i \prec_\rho j \Leftrightarrow \rho(i) \prec \rho(j)$. A cell C_i is said to precede another cell C_j if $i \prec_\rho j$. To each cell C_i, $i \in I$ we associate a reference point $\xi_i \in C_i$. Two cells C_j and C_i, $i \neq j$ are said to be neighbour cells if $\|\xi_i - \xi_j\| \leq R_P$, where $R_P > 0$ is the POMM interaction range. Let n_i denote the number of points in cell C_i. For all $i \in I$, let $s_{i,R_P} = \sum_{j \in I : j \prec_\rho i} n_j \mathbf{1}[\|\xi_j - \xi_i\| \leq R_P]$ be the number of points in the preceding neighbour cells of C_i. Note that we have suppressed the dependence on ρ and $\{\xi_i : i \in I\}$ in the notation s_{i,R_P}.

Now assume that n_i conditional on $\{n_j : j \prec_\rho i\}$ is Poisson distributed with mean $\lambda_i = \beta_P |C_i| \gamma_P^{s_{i,R_P}}$, where $\beta_P > 0$ and $0 \leq \gamma_P \leq 1$ are parameters. Then

$$\mathbb{P}(\{n_i : i \in I\}) = \prod_{i \in I} e^{-\lambda_i} \lambda_i^{n_i} / n_i!.$$

Conditional on $\{n_i : i \in I\}$ let X_i, $i \in I$ be independent point processes where X_i is a binomial point process on C_i conditioned to have n_i points. Then $X = \cup_{i \in I} X_i$ is a point process with normalised density

$$f_P(x|\beta_P, \gamma_P, R_P, \rho) = \exp(-\beta_P \sum_{i \in I} |C_i| \gamma_P^{s_{i,R_P}(x)}) \beta_P^{n(x)} \prod_{i \in I} \gamma_P^{n_i(x) s_{i,R_P}(x)} \quad (5.6)$$

w.r.t. μ, where $n_i(x) = n(x \cap C_i)$ and $s_{i,R_P}(x) = \sum_{j \in I: j \prec_\rho i} n_j(x) \mathbf{1}[\|\xi_j - \xi_i\| \leq R_P]$. On the left hand side of (5.6) we have suppressed the dependence on $\{C_i : i \in I\}$ and $\{\xi_i : i \in I\}$ as they are usually kept fixed. See Cressie et al. (2000) for more details on how (5.6) is derived.

In general we will not keep the permutation ρ fixed. This is in an attempt to reduce any order dependent bias. Not expecting a relation between ρ and (β, γ, R), we introduce ρ into the general setup as an additional auxiliary variable. This is done by replacing $f(x|y, \theta)$ in the numerator of (5.2) by $f(x|y, \theta, \rho)\tilde{f}(\rho|y, \theta)$ and similarly in the denominator. Assuming that ρ is uniformly distributed over all permutations of I independent of y and θ imply that $\tilde{f}$ cancels in (5.2).

It remains to specify (β_P, γ_P, R_P) given the parameters $\theta = (\beta, \gamma, R)$ for a Strauss process. Let $(\beta_P, \gamma_P, R_P) = g(\beta, \gamma, R) \equiv (g_1(\theta), g_2(\theta), g_3(\theta))$ where $g : [0, \infty) \times (0, 1] \times [0, \infty) \mapsto [0, \infty) \times (0, 1] \times [0, \infty)$. Then the POMM auxiliary point process specified by ρ and the Strauss parameter θ has density

$$f(x|(\beta, \gamma, R), \rho, y) = f_P(x|g(\beta, \gamma, R), \rho)$$

w.r.t. μ.

When specifying g we note that $\sum_{i \in I} s_{i,R_P}(x)$ tends to $s_{R_P}(x)$ as $\text{card}(I) \to \infty$ under the condition that $\max_{i \in I} |C_i|$ tends to zero, where $\text{card}(I)$ is the cardinality of I. This motivates setting $g_3(\theta) = R$. Ideally we would like that $(g_1(\theta), g_2(\theta)) = \mathbb{E}[\text{argmax}_{(\tilde{\beta}, \tilde{\gamma})} \, f_P(X|\tilde{\beta}, \tilde{\gamma}, R, \rho)]$ where X is a Strauss process specified by θ and ρ is uniformly distributed. In general we have no theoretical expression for this expectation and hence we have to approximate it. In Section 5.2.3 we elaborate on how g is approximated. In addition we have considered the identity mapping $g(\beta, \gamma, R) = (\beta, \gamma, R)$.

The simulation study in Section 5.2.2 shows that the POMM auxiliary point process works better than the fixed MLE Strauss point process but only when the cell size is small compared to the interaction range R. The high number of cells thus needed imply that evaluating (5.6) is much more time consuming than evaluating (5.5).

5.2.2 Simulation study

In this section we compare the three auxiliary point processes considered in Section 5.2.1 when the likelihood is given by (5.3). Details on the Metropolis-Hastings algorithms used for the simulations in this section are given in Section 5.2.3. The "data" y in Figure 5.1 is a realisation of a Strauss point process on the unit square $S = [0, 1]^2$ with $(\beta, \gamma, R) = (100, 0.5, 0.05)$ generated using dominated CFTP (Kendall & Møller 2000). Assuming $R = 0.05$ fixed we have sufficient statistics $n(y) = 75$ and $s_R(y) = 10$. A priori we assume that β and γ are independent and uniformly distributed on $[0, 150]$ and $(0, 1]$, respectively. Using the grid approach the MLE is $(\hat{\beta}, \hat{\gamma}) = (108, 0.4)$.

For the POMM point process we divide S into $N \times N$ square cells of equal size. When choosing N one should take into account that the side length of the

Aux.proc.	g	σ_β	σ_γ	MAcP	Extr	c_β	c_γ
Fixed Poisson		2	0.05	0.128	0.151	0.88	0.53
Fixed Strauss		2	0.05	0.393	0.031	0.79	0.46
POMM (N=50)	MLE	2	0.05	0.246	0.055	0.85	0.46
POMM (N=100)	MLE	2	0.05	0.321	0.013	0.79	0.38
POMM (N=100)	MLE	4	0.1	0.298	0.030	0.52	0.21
POMM (N=100)	(β, γ)	2	0.05	0.17	0.31	0.86	0.54
POMM (N=200)	MLE	2	0.05	0.406	0.002	0.75	0.33

Table 5.1: Empirical results. Aux.Proc.: Type of auxiliary process. g: Type of mapping used in case of a POMM point process. σ_β and σ_γ: Parameters for the proposal distribution for β' and γ', see Section 5.2.3 for more details. MAcP: Mean acceptance probability. Extr: Fraction of Hastings ratios below $\exp(-10)$. c_β and c_γ: Lag 100 autocorrelation for β and γ, respectively.

square cells is $1/N$ and the interaction range is $R = 0.05$. We have chosen the values $N = 50, 100, 200$.

For each auxiliary process considered we have generated one million updates using the appropriate Metropolis-Hastings algorithm described in Section 5.2.3. As the posterior distribution of (β, γ) is the same no matter which auxiliary process is considered the main difference is how well the Metropolis-Hastings chains are mixing. As indicators for the mixing properties, we consider the lag 100 autocorrelation of β and γ and the mean acceptance probability for the Metropolis-Hastings algorithm. Further, we consider the fraction of Hastings ratios (5.2) below $\exp(-10)$. This is because early experiments with the algorithm showed that trace-plots of $n(x)$ and $s_R(x)$ may exhibit seemingly satisfactory mixing properties for several million updates and then get stuck — sometimes for more than 100,000 updates. The results for the different auxiliary processes are summarised in Table 5.1.

Comparing results for the fixed Poisson and the fixed Strauss processes the most striking difference is in the fraction of extremely low Hastings ratios. The high fraction for the fixed Poisson process indicates a high risk of getting stuck with poor mixing as a result. This is confirmed by trace plots of $n(x)$ and $s_R(x)$ (not shown here) as well as the empirical posterior distributions (not shown here) which do not look as smooth as those for the POMM point process with $N = 200$ shown in Figure 5.2.

The results in Table 5.1 show that using the POMM point process with $N = 50$ as the auxiliary process the Metropolis-Hastings chain is mixing worse than using the fixed Strauss process. Using $N = 100$ improves all statistics compared to $N = 50$. Further, the case $N = 100$ seems to be a slight improvement over the fixed Strauss process, but with a lower mean acceptance probability. For $N = 200$ the mean acceptance probability is close to that of the fixed Strauss process. But much fewer extremely low Hastings ratios indicates that the POMM point process is a better approximation. Figure 5.2 shows the

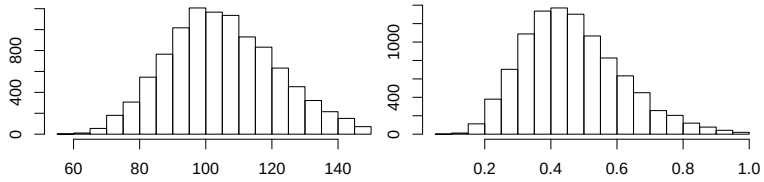

Figure 5.2: Empirical posterior distribution of β (left plot) and γ (right plot) generated using a POMM auxiliary process with $N = 200$.

posterior distribution for β and γ when the auxiliary process is the POMM with $N = 200$. As we have used uniform priors the posterior maximum is equivalent to the MLE. Hence we can confirm that $\hat{\gamma} = 0.4$ is not far from the posterior mode whereas $\hat{\beta} = 108$ is around 10% higher.

As an additional experiment we have used the POMM point process with $N = 100$ together with a proposal distribution with higher variance than above. This results in improved autocorrelation and slightly lower mean accept probability but a higher fraction of extremely low Hastings ratios. In the case $N = 100$ we have also tried letting g be the identity mapping avoiding the initial approximation of g. The resulting mean acceptance probability is rather low — only for the fixed Poisson is it lower.

Comparing computer times using the POMM with $N = 100$ is not much slower than using the fixed Strauss process. Using $N = 200$ takes twice as long as $N = 100$.

In conclusion it seems that a cell side length less than $R/10$ is needed to obtain a significant improvement using a POMM auxiliary process compared to using a fixed Strauss process.

5.2.3 Simulation details

We now give some details on the Metropolis-Hastings algorithms used for simulating the posterior using the auxiliary processes considered in Sections 5.2.1 and 5.2.2. Further, we give some details on the approximation of the mapping g.

In the Metroplis-Hastings algorithms for fixed Poisson and fixed Strauss updates are done as follows with f in step 3 replaced by (5.4) or (5.5), respectively. As initial values we choose $\theta = (n(y), 1, 0.05)$ and x is a realisation of a Poisson point process on S with intensity β.

Then, if $\theta = (\beta, \gamma, R)$ and x comprise the current state of the Metropolis-Hastings algorithm we update as follows.

1. Generate $\beta' \sim \mathcal{N}(\beta, \sigma_\beta)$ and $\gamma' \sim \mathcal{N}(\gamma, \sigma_\gamma)$ and set $\theta' = (\beta', \gamma', R)$.
2. Generate a realisation x' from a Strauss process specified by θ' using dominated CFTP.

3. With probability

$$\min\Big\{1, \mathbf{1}[0 \le \beta' \le 150, 0 < \gamma' \le 1] \times \left(\frac{\beta'}{\beta}\right)^{n(y)} \left(\frac{\gamma'}{\gamma}\right)^{s_R(y)} \frac{f(x'|y,\theta')}{f(x|y,\theta)} \frac{\beta^{n(x)}\gamma^{s_R(x)}}{\beta'^{n(x')}\gamma'^{s_R(x')}}\Big\}$$

set $\theta = \theta'$ and $x = x'$ otherwise do nothing.

Here $\mathcal{N}(\nu, \sigma)$ denotes a one-dimensional normal distribution with mean ν and variance σ^2. For both the fixed Poisson process and the fixed Strauss process we have used $\sigma_\beta = 2$ and $\sigma_\gamma = 0.05$.

As the POMM approach includes an extra auxiliary variable, the random permutation ρ, an additional step is required in the update above. Assume that the current state is made up of $\theta = (\beta, \gamma, R)$, ρ and x then an update consists of steps 1 and 2 above followed by

3. Generate uniform random permutation ρ'.

4. With probability

$$\min\Big\{1, \mathbf{1}[0 \le \beta' \le 150, 0 < \gamma' \le 1] \times \left(\frac{\beta'}{\beta}\right)^{n(y)} \left(\frac{\gamma'}{\gamma}\right)^{s_R(y)} \frac{f(x'|y,\theta',\rho')}{f(x|y,\theta,\rho)} \frac{\beta^{n(x)}\gamma^{s_R(x)}}{\beta'^{n(x')}\gamma'^{s_R(x')}}\Big\}$$

set $\theta = \theta'$, $\rho = \rho'$ and $x = x'$ otherwise do nothing.

In general we have used $\sigma_\beta = 2$ and $\sigma_\gamma = 0.05$ except for one experiment with $N = 100$ where $\sigma_\beta = 4$ and $\sigma_\gamma = 0.1$.

When the mapping g is not the identity it is approximated as follows. Based on the range of the empirical posterior distributions in the fixed Strauss case (not shown here) we define a grid $G = \{50, 52, \ldots, 150\} \times \{0.1, 0.2, \ldots, 1.0\} \times \{0.05\}$. For each grid point $(\beta, \gamma, R) \in G$ we generate 10 independent realisations $x_1, \ldots, x_{10}$ of a Strauss point process with parameters (β, γ, R) (using dominated CFTP) together with the generation of 10 independent random permutations $\rho_1, \ldots, \rho_{10}$. Then $g(\beta, \gamma, R)$ is approximated by

$$(g_1(\theta), g_2(\theta)) = \frac{1}{10}\sum_{i=1}^{10} \text{argmax}_{(\tilde{\beta},\tilde{\gamma})} f(x_i|\tilde{\beta}, \tilde{\gamma}, R, \rho_i), \quad \text{and } g_3(\theta) = R.$$

For $(\beta, \gamma, 0.05) \notin G$ we set $g(\beta, \gamma, 0.05) = g(\tilde{\beta}, \tilde{\gamma}, 0.05)$ where $(\tilde{\beta}, \tilde{\gamma}, 0.05) \in G$ is the grid point closest to $(\beta, \gamma, 0.05)$.

Figure 5.3 shows $g_1(\beta, \gamma, R) - \beta$ and $g_2(\beta, \gamma, R) - \gamma$ for a range of β and γ values when $N = 200$. Results for $N = 50$ and $N = 100$ are almost identical to those for $N = 200$. The difference $g_1(\beta, \gamma, R) - \beta$ is significant in cases of strong interaction, i.e. for combinations of low values of γ and high values of β.

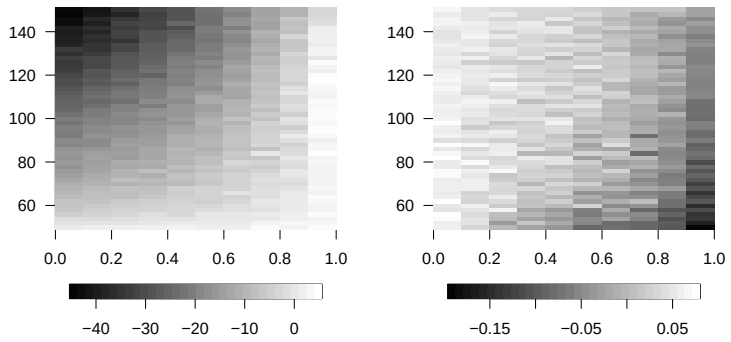

Figure 5.3: Plot of difference between $g(\theta)$ and θ for $\theta \in G$. Left plot: $g_1(\beta, \gamma, R) - \beta$. Right plot: $g_2(\beta, \gamma, R) - \gamma$.

This could be explained by the fact that the interaction in the POMM auxiliary process is weaker than in the Strauss process with same values of β, γ and R. In general the difference $g_2(\beta, \gamma, R) - \gamma$ is small except for cases of little (or no) interaction, i.e. when β is small and γ is close to (or equal) 1.

Bibliography

Besag, J. (1975), 'Statistical analysis of non-lattice data', *The Statistician* **25**, 179–195.

Cressie, N. & Davidson, J. L. (1998), 'Image analysis with partially ordered markov models', *Comput. Statist. Data Anal.* **29**, 1–26.

Cressie, N., Zhu, J., Baddeley, A. J. & Nair, M. G. (2000), 'Directed markov point processes as limites of partially ordered markov models', *Methodol. Comput. Appl. Probab.* **2**, 5–21.

Davidson, J. L., Cressie, N. & Hua, X. (1999), 'Texture synthesis and pattern recognition for partially ordered markov models', *Pattern Recognition* **32**, 1475–1505.

Gelman, A. & Meng, X.-L. (1998), 'Simulating normalizing constants: from importance sampling to bridge sampling to path sampling', *Statist. Sci.* **13**, 163–185.

Geyer, C. J. & Thompson, E. A. (1992), 'Constrained Monte Carlo maximum likelihood for dependant data', *Journal of the Royal Society of Statistics* **B 54**.

Kelly, F. P. & Ripley, B. D. (1976), 'A note on Strauss's model for clustering', *Biometrika* **63**, 357–360.

Kendall, W. S. & Møller, J. (2000), 'Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes', *Adv. Appl. Prob.* **32**, 844–865.

Meng, X.-L. & Wong, W. H. (1996), 'Simulating ratios of normalizing constants via a simple identity: a theoretical exploration', *Statist. Sinica* **6**.

Penttinen, A. (1984), *Modelling Interaction in Spatial Point Patterns: Parameter Estimation by the Maximum Likelihood Method*, Number 7 in Jyväskylä Studies in Computer Science, Economics, and Statistics.

Ripley, B. D. (1988), *Statistical Inference for Spatial Processes*, Cambridge University Press, Cambridge.

Strauss, D. J. (1975), 'A model for clustering', *Biometrika* **62**, 467–475.

6 Non-parametric Bayesian inference for inhomogeneous pairwise interaction point processes

Kasper K. Berthelsen and Jesper Møller

Department of Mathematical Sciences, Aalborg University,
Fredrik Bajers Vej 7G, DK-9220 Aalborg, Denmark

Abstract. We assume that a inhomogeneous point pattern can be modelled by a pairwise interaction point process where the inhomogeneity is model by an location dependent first order interaction term. In a Bayesian setup we assume a priori that the first order term is a shot noise process and the second order term is of a parametric nature. Simulation of the posterior is done using an auxiliary variable technique which avoids the unknown normalising constant in the likelihood. The auxiliary variable used is an example of a partially ordered Markov model. The method is applied to a real data set.

6.1 Introduction

Observed spatial point patterns often show signs of inhomogeneous properties. Such observations can be modelled by inhomogeneous point processes. A number of models for inhomogeneous point processes have been proposed. The most basic one is an inhomogeneous Poisson process. Heikkinen & Arjas (1998) consider a non-parametric Bayesian analysis of the intensity function for an inhomogeneous Poisson process. They assume a priori that the intensity function is piecewise constant and use the posterior mean as their estimate. In many applications it seems natural to expect some degree of interaction between points where the first or the second order interaction term is location dependent. Jensen & Nielsen (2000) consider modelling an inhomogeneous point process by transforming a homogeneous pairwise interaction process. Hahn, Jensen, van Lieshout & Nielsen (2003) model an inhomogeneous point process by local scaling which in effect means that the interaction range is location dependent. In this paper we consider the situation where the inhomogeneous point process is model by a pairwise interaction point process with an inhomogeneous first order term. More specific we consider inhomogeneous pairwise interaction point

processes on a bounded region $W \subset \mathbb{R}^2$ with normalised density

$$\pi(x|\beta,\varphi) = \frac{1}{Z_{(\beta,\varphi)}} \prod_{\xi\in x} \beta(\xi) \prod_{\{\xi,\eta\}\subseteq x:\xi\neq\eta} \varphi(\|\xi-\eta\|) \tag{6.1}$$

w.r.t. unit rate Poisson on W. Here $Z_{(\beta,\varphi)}$ is the normalising constant, $\beta : \mathbb{R}^2 \mapsto [0,\infty)$ is a locally integrable intensity function, and $\varphi : [0,\infty) \mapsto [0,1]$ is a repulsive interaction function. The normalising constant is in general unknown which results in some difficulties considered in Section 6.3.

For specificity we choose the following interaction function

$$\varphi(r) = \begin{cases} 0 & \text{if } r \le R_0 \\ 1 - \left(1 - \left(\frac{r-R_0}{R-R_0}\right)^m\right)^m & \text{if } R_0 < r \le R \\ 1 & \text{if } r > R. \end{cases} \tag{6.2}$$

Defining the interaction range as $\sup\{r \ge 0 : \varphi(r) < 1\}$ the parameter $R \ge 0$ is equivalent to the interaction range for (6.2) and $0 \le R_0 < R$ is a hard core distance. Further, for $m > 1$ the interaction function (6.2) is smooth in the sense that it is differentiable w.r.t. r. Setting $m = 2$ and $R_0 = 0$ we get that (6.2) is one of the interaction functions considered by Diggle, Fiksel, Grabarnik, Ogata, Stoyan & Tanemura (1994).

This paper is organised as follows. In Section 6.2 we specify the prior distributions for the intensity function β and the parameters of the interaction function and derive the posterior distribution. Section 6.3 contains a description of how we sample from the posterior using an auxiliary variable technique, further we consider the specific auxiliary process used in the analysis of a real data set in Section 6.4.

6.2 Prior and posterior

Assume that the intensity function β can be approximated by

$$\beta(\xi;\psi) = \sum_j \lambda_j \mathcal{K}(\xi;c_j) \tag{6.3}$$

where $\psi = \{(c_1,\lambda_1),(c_2,\lambda_2),\ldots\}$, $(c_j,\lambda_j) \in \mathbb{R}^2\times[0,\infty)$ and $\mathcal{K}(\cdot;\cdot)$ is a kernel, i.e. $\mathcal{K}(\cdot;c)$ is a probability density for a 2-dimensional random variable for all $c \in \mathbb{R}^2$. For specificity we assume that $\mathcal{K}(\cdot;c)$ is the normalised density of a bivariate normal distributed random variable with mean c and dispersion matrix $\sigma^2 I$, $\sigma > 0$. This choice justifies that we in the following sometimes refer to (c,λ) as a kernel centred at c (and scaled by λ).

We assume a priori that ψ is a marked Poisson process where $\mathcal{C} = \{c_1, c_2, \ldots\}$ is a homogeneous Poisson process on $\mathbb{R}^2$ with intensity κ and that conditional on $\mathcal{C}$ the marks λ_j are i.i.d. gamma distributed with scale parameter $\alpha_1 > 0$ and shape parameter $\alpha_2 > 0$. The latter we write as $\lambda_j \sim \text{gamma}(\alpha_1,\alpha_2)$.

This specification of ψ is equivalent to letting ψ be an inhomogeneous Poisson process on $\mathbb{R}^2 \times [0,\infty)$ with intensity function $\kappa\alpha_1^{-\alpha_2}\lambda^{\alpha_2-1}\exp(-\lambda/\alpha_1)/\,\Gamma(\alpha_2)$, i.e. proportional to the density of a gamma distribution with scale parameter α_1 and shape parameter α_2.

Assume for a moment that $\varphi \equiv 1$ which is equivalent to no interaction. Then a point process which conditional on a realisation of the random first order term (6.3) has density (6.1) w.r.t. a unit rate Poisson process is a so-called shot noise Cox process (SNCP). For more details on SNCPs, see e.g. Møller (2003). Assume instead that φ is a general repulsive interaction function. The resulting point process can then be thought of as a generalised SNCP.

Using results in Møller (2003) for SNCPs we obtain the prior mean $\mathbb{E}\,(\beta(\xi;\psi)) = \kappa\alpha_1\alpha_2$ and prior variance $\mathbb{V}\mathrm{ar}(\beta(\xi;\psi)) = \kappa(\alpha_1\alpha_2)^2/(4\pi\sigma^2)$.

To make the approximation (6.3) feasible the infinite sum is restricted to a sum over marked points (c_i,λ_i) with $c_i \in W_{\mathrm{ext}}$ where $W_{\mathrm{ext}} \supseteq W$ is a bounded set. Redefining $\psi = \{(c_i,\lambda_i) : c_i \in \mathcal{C} \cap W_{\mathrm{ext}}\}$ we have that ψ is almost surely finite. Let ν be an inhomogeneous Poisson process on $W_{\mathrm{ext}} \times [0,\infty)$ with intensity measure

$$\rho(F) = \int_{[0,\infty)} \int_{W_{\mathrm{ext}}} \mathbf{1}[(c,\lambda) \in F]\alpha_1^{-\alpha_2}\lambda^{\alpha_2-1}\exp(-\lambda/\alpha_1)/\Gamma(\alpha_2)\mathrm{d}c\mathrm{d}\lambda$$

for all $F \subseteq W_{\mathrm{ext}} \times [0,\infty)$. Then ψ is an inhomogeneous Poisson process with density $\kappa^{n(\psi)}$ w.r.t. ν where $n(\psi)$ denotes the cardinality of ψ.

The restriction of ψ introduces an edge effect on the prior distribution of the first order term due to the missing contributions from kernels centred outside W_{ext}. To asses the degree of edge effect introduced by the restriction of ψ define

$$M = \int_W \sum_{j:c_j \in \mathcal{C}\setminus W_{\mathrm{ext}}} \lambda_j \mathcal{K}(\xi;c_j)\mathrm{d}\xi.$$

Here M can be interpreted as the total contribution from kernels centred outside W_{ext} missed when restricting ψ. Going along similar lines as in Møller (2003) we have

$$\mathbb{E}M = \kappa\alpha_1\alpha_2 \int_{\mathbb{R}^2\setminus W_{\mathrm{ext}}} \int_W \mathcal{K}(\xi;c)\mathrm{d}\xi\mathrm{d}c \le \kappa\alpha_1\alpha_2 \int_{\mathbb{R}^2\setminus W_{\mathrm{ext}}} \int_W k(\xi,c)\mathrm{d}\xi\mathrm{d}c, \quad (6.4)$$

where $k : \mathbb{R}^2 \times \mathbb{R}^2 \mapsto \mathbb{R}$ is chosen so that the integration becomes easier and $k(\xi;c) \ge \mathcal{K}(\xi;c)$ for all $(\xi,c) \in W \times (\mathbb{R}^2\setminus W_{\mathrm{ext}})$.

Assume that W is a rectangular region with side lengths a and b and let $W_{\mathrm{ext}} = \{\xi \in \mathbb{R}^2 : \delta(\xi,W) \le \Delta\}$ for $\Delta \ge 0$ where $\delta(A,B) = \inf\{\|\xi-\eta\| : \xi \in A, \eta \in B\}$ for any subsets $A, B \subseteq \mathbb{R}^2$. Then $k(\xi,c) = (2\pi\sigma^2)^{-1}e^{-\delta(c,W)^2/(2\sigma^2)} \ge \mathcal{K}(\xi,c)$ for all $(\xi,c) \in W \times (\mathbb{R}^2\setminus W_{\mathrm{ext}})$ and

$$\mathbb{E}M \le \kappa\alpha_1\alpha_2|W| \int_\Delta^\infty \left[2(a+b)/(2\pi\sigma^2) + r/\sigma^2\right] e^{-r^2/(2\sigma^2)}\mathrm{d}r. \quad (6.5)$$

Typically values for κ, α_1, α_2, a, b and σ^2 are given together with an upper bound on $\mathbb{E}M$. In this case (6.5) can be solved numerically to give a lower bound on Δ.

For simplicity we assume a priori that the three parameters R, R_0 and m in the interaction function (6.2) are independent and uniformly distributed on $[b_{R_0}, b_R]$, $[b_{R_0}, b_R]$ and $[b_{m_1}, b_{m_2}]$, respectively, with the restriction that $R_0 < R$ and $b_{R_0}, b_R, b_{m_1}, b_{m_2} \in [0, \infty)$, $b_{m_1} < b_{m_2}$.

Under the above assumptions the posterior density of $\theta = (\psi, R, R_0, m)$ is

$$\pi(\psi, R, R_0, m|y) \propto \kappa^{n(\psi)} \frac{1}{Z_\theta} \prod_{\xi \in y} \beta(\xi; \psi) \prod_{\{\xi,\eta\} \subseteq y: \xi \neq \eta} \varphi(\|\xi - \eta\|) \times \mathbf{1}\big[R_0 < R, R \in [b_{R_0}, b_R], R_0 \in [b_{R_0}, b_R], m \in [b_{m_1}, b_{m_2}]\big] \quad (6.6)$$

w.r.t. the product measure of ν and the Lebesgue measure on $[b_{R_0}, b_R] \times [b_{R_0}, b_R] \times [b_{m_1}, b_{m_2}]$.

6.3 Sampling from the posterior

Sampling θ from the posterior distribution (6.6) using a Metropolis-Hastings algorithm in the "conventional" way involves evaluating ratios of the unknown normalising constant Z_θ. There exists a number of methods for approximating such ratios, e.g. importance, bridge and path sampling, see Gelman & Meng (1998) for more details on these methods and the relation between them. In this section we consider an alternative to these methods introduced by Berthelsen, Møller, Pettitt & Reeves (2003) which avoids the unknown normalising constant by introducing an auxiliary variable.

The idea is as follows. Introduce an auxiliary point process x with normalised density $f(\cdot|\theta, y)$ defined on the same space as the likelihood $\pi(\cdot|\theta)$. Then instead of sampling from the posterior $\pi(\theta|y) = \pi(y|\theta)\pi(\theta)$ we sample (x, θ) from the joint density $\pi(x, \theta|y) = f(x|\theta, y)\pi(y|\theta)\pi(\theta)$. Note that the marginal distribution of θ is given by $\pi(\theta|y)$. Sampling (x, θ) using a Metropolis-Hastings algorithm we assume that the proposal distribution for (x', θ') is of the form

$$p(x', \theta'|x, \theta, y) = \pi(x'|\theta')p(\theta'|\theta, y)$$

where $p(\cdot|\theta, y)$ is assumed to be a normalised density for all θ. A proposal can then be generated by first generating θ' from $p(\cdot|\theta, y)$ and then x' from $\pi(\cdot|\theta')$. We then arrive at the Hastings ratio

$$H(x', \theta'|x, \theta, y) = \frac{f(x'|\theta', y)\pi(y|\theta')\pi(\theta')\pi(x|\theta)p(\theta|\theta', y)}{f(x|\theta, y)\pi(y|\theta)\pi(\theta)\pi(x'|\theta')p(\theta'|\theta, y)} \quad (6.7)$$

where the unknown normalising constants cancel. In practice the success of this method relies on how well the auxiliary process approximates $\pi(\cdot|\theta)$.

Using the approach in Berthelsen et al. (2003) we propose to use a partially ordered Markov model (POMM) point process as the auxiliary point process. It has the advantages that it incorporates some degree of interaction and still has a

known normalising constant. POMMs were introduced by Cressie & Davidson (1998) and Davidson et al. (1999) in the analysis of gray scaled images. A POMM point process X is simulated as follows. Initially W is divided into N disjoint subsets $\{C_i \subset W : i = 1, \ldots, N\}$, $\cup_{i=1}^{N} C_i = W$, $C_i \cap C_j = \emptyset$ for $i \neq j$. Sequentially for $i = 1, \ldots, N$ generate a Poisson distributed random variable n_i with mean λ_i and conditional on n_i let $X \cap C_i$ be a realisation of a binomial point process conditioned to have n_i points. The key feature of POMM point processes is that λ_i, $i = 2, \ldots, N$, is allowed to depend on $X \cap C_j$, $j = 1, \ldots, i-1$ whereby interaction can be introduced.

In the present setting we assume that W is rectangular and is divided into rectangular subsets of equal size and shape. Berthelsen et al. (2003) assign the indices to the subsets randomly. For computational reasons we choose to index the subsets systematically row-wise. To each subset C_i, $i = 1, \ldots, N$, we associate a reference point $\xi_i \in C_i$ and let

$$\lambda_i = |C_i| \omega \beta(\xi_i; \psi) \Phi_i(x), \tag{6.8}$$

where $\Phi_i(x) = \prod_{j<i} \prod_{\eta \in (x \cap C_j)} \varphi(\|\eta - \xi_i\|)$ and $\omega > 0$. Note that compared to Berthelsen et al. (2003) distances in $\Phi_i(x)$ are between the reference point ξ_i associated with C_i and points in $x \cap C_j$, $j < i$, and not between pairs of reference points. It has not been investigated how much this refinement improves the POMM point process as an approximation of a pairwise interaction point process.

The parameter ω in (6.8) is introduced because Berthelsen et al. (2003) find that the empirical mean MLE of β for the POMM point process is usually lower than β for the corresponding true pairwise interaction point process. Specifying the POMM point process ω should depend on the degree of interaction, e.g. the magnitude of R with the restriction that $\omega = 1$ in the Poisson case $R = 0$. On the other hand, if ω is kept fixed a simulation study in Berthelsen et al. (2003) suggests that this will result in poor mixing properties for the resulting Metropolis-Hastings algorithm.

With the above specification of the POMM point process it has density

$$f(x|\psi, R, R_0, m) = \exp\left[-\sum_{i=1}^{N} \omega |C_i| \beta(\xi_i; \psi) \Phi_i(x)\right] \prod_{i=1}^{N} [\omega \beta(\xi_i; \psi) \Phi_i(x)]^{n_i(x)} \tag{6.9}$$

w.r.t. a unit rate Poisson process on W.

One problem with the POMM point process is the lack of interaction between points located in the same subset. To illustrate this problem assume that the current state (x, θ) has at most one point in each subset C_i, $i = 1, \ldots, N$, and we propose a new state (x', θ') where x' contains two points in the same subset separated by a distance less than 10^{-3}. If the interaction function is given by (6.2) with $m = 2$, $R_0 = 0$ and $R = 0.1$ we have $\varphi(10^{-3}) \approx 2 \times 10^{-4}$. Due to the lack of interaction between points located in the same subset a term of this magnitude is "missed" in the Hastings ratio. This in turn implies a high probability of accepting the proposal and a risk of relatively low acceptance probabilities for successive proposals.

When generating proposal of (x, θ) we need a point process sample from $\pi(\cdot|\theta)$. This is done using the dominated coupling from the past (dominated CFTP) algorithm for perfect simulation given by Kendall & Møller (2000).

When sampling from the posterior we use a hybrid Metropolis-Hastings algorithm with $k = 6$ types of updates. Define $p = (p_1, p_2, \ldots, p_k)$ where $p_i \geq 0$ and $\sum_{i=1}^{k} p_i = 1$. Then each update consist of doing the ith type of update with probability p_i, $i \in \{1, \ldots, k\}$.

Assume the current state is given by (x, θ) with $\theta = (\psi, R, R_0, m)$. All updates consists in proposing a θ', generating x' from $\pi(x'|\theta')$ and accepting the proposal (x', θ') with probability $\min\{1, H(x', \theta'|x, \theta, y)\}$ where H is given by (6.7). Only the proposal distribution differs for the different updates.

With probability p_1 propose to add a kernel (c', λ') where c' is uniformly distributed on W_{ext} and $\lambda' \sim \text{gamma}(\alpha_1, \alpha_2)$, i.e. the proposal distribution of (c', λ') has density $|W_{\text{ext}}|^{-1}$ w.r.t. ρ. The acceptance probability is then calculated using $\theta' = (\psi \cup \{(c', \lambda')\}, R, R_0, m)$, $p(\theta'|\theta, y) = |W_{\text{ext}}|^{-1}$ and $p(\theta|\theta', y) = (1 + n(\psi))^{-1}$.

With probability p_2 propose to remove a kernel (c', λ') uniformly randomly selected from ψ. The acceptance probability is then calculated with $\theta' = (\psi \backslash \{(c', \lambda')\}, R, R_0, m)$, $p(\theta'|\theta, y) = n(\psi)^{-1}$ and $p(\theta|\theta', y) = |W_{\text{ext}}|^{-1}$.

With probability p_3 we propose to translate a kernel (c, λ) uniformly selected from ψ by a distance c'. The translation c' follows a bivariate normal distribution with mean zero and dispersion matrix $\sigma_m^2 I$. The acceptance probability is then calculated using $\theta' = (\psi \backslash \{(c, \lambda)\} \cup \{(c + c', \lambda)\}, R, R_0, m)$ if $c + c' \in W_{\text{ext}}$ otherwise the proposal is rejected. Further, note that the proposal distributions cancel due to symmetry.

With probabilities p_4, p_5 and p_6 we propose to update R, R_0 and m, respectively. In all cases the proposal is uniformly distributed on the interval centred at the current value with width $2\epsilon_R$, $2\epsilon_{R_0}$ and $2\epsilon_m$, respectively. Calculating the acceptance probability θ' is specified in an obvious way and proposal distributions cancel due to symmetry.

Note that the addition and removal of kernels is essentially an application of the spatial birth-and-death algorithm given by Geyer & Møller (1994) which in turn is a special case of the reversible jump algorithm (Green 1995).

6.4 Data analysis

The data shown in the left plot of Figure 6.1 is the location of cells in a section of the mocous membrane of the stomach of a healthy rat. The left hand side of the plot corresponds to where the stomach cavity begins and the right hand side corresponds to where the muscle tissue begins. The data consists of 617 points rescaled to a window $W = [0, 1] \times [0, 0.893]$. The right plot in Figure 6.1 is an estimate of the cell intensity. This estimate is given by the empirical posterior mean for $\beta(\cdot; \psi)$ based on samples generated using the simulation scheme above with $\varphi \equiv 1$ and $\omega \equiv 1$ fixed and skipping updates of R_0, R and m.

The summary statistics in the upper row of plots in Figure 6.2 suggests that

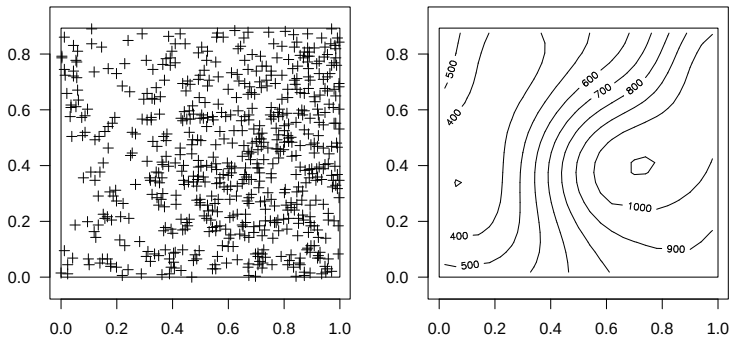

Figure 6.1: Left plot: locations of 617 cells in a 2D section of the mocous membrane of the stomach of a healthy rat. Right plot: estimate of the cell intensity.

data is not a realisation of an inhomogeneous Poisson process but show sign of small scale inhibition for inter point distances less than 0.015.

The cell data was originally analysed by Nielsen (2000) who models the data by a transformation of a Strauss point process. A Strauss point process on W has density of the form (6.1) where $\beta(\xi) \equiv \beta$ is constant and $\varphi(r) = \gamma^{\mathbf{1}[r \leq R]}$ with $0 < \gamma \leq 1$ and $R > 0$. Nielsen (2000) finds that the transformed Strauss point process fit the data well with parameters $\beta = 760$, $\gamma = 0.09$ and $R = 0.007$. This leads us to choose κ, α_1 and α_2 so that $\kappa\alpha_1\alpha_2 \approx 760$. Recalling that we are at most adding or removing a single kernel in each update a high mean number of kernels would result in slow mixing. For these reasons we have chosen $\kappa = 50$, $\alpha_1 = 15$ and $\alpha_2 = 1$. Furthermore, we propose to use $\sigma = 0.15$.

Assume an upper bound $\mathbb{E}M \leq \kappa\alpha_1\alpha_2|W|0.001$ then solving (6.5) numerically gives $\Delta \geq 0.59$. For simplicity we set $W_{\text{ext}} = [-0.59, 1.59] \times [-0.59, 1.483]$ which includes the extended region defined for (6.5).

Berthelsen et al. (2003) conclude that the POMM point process is a good approximation of a pairwise interaction point process when the side lengths of the rectangular subsets are less than one 10th of the interaction range. Expecting the interaction range to be larger than 0.01 in general we divide W into $N = 1000 \times 1000$ rectangular subsets of equal size and shape. The problem of two points in the same subset is then avoided by setting $b_{R_0} = 1.34 \times 10^{-3}$. For comparison the shortest inter point distance in the data is 5.93×10^{-3}.

Specifying ω we use that the MLE of ω is

$$\hat{\omega} = \frac{\sum_{i=1}^{N} n_i(x)}{\sum_{i=1}^{N} \beta(\xi_i; \psi)|C_i|\Phi_i(x)}.$$

For a POMM point processes with α_1, α_2 and κ as above and $R = 0.015$, $R_0 = 0$ and $m = 2$ fixed the empirical mean MLE for ω was 0.89. Combining this with the fact that $R = 0$ should imply $\omega = 1$ we have used the following

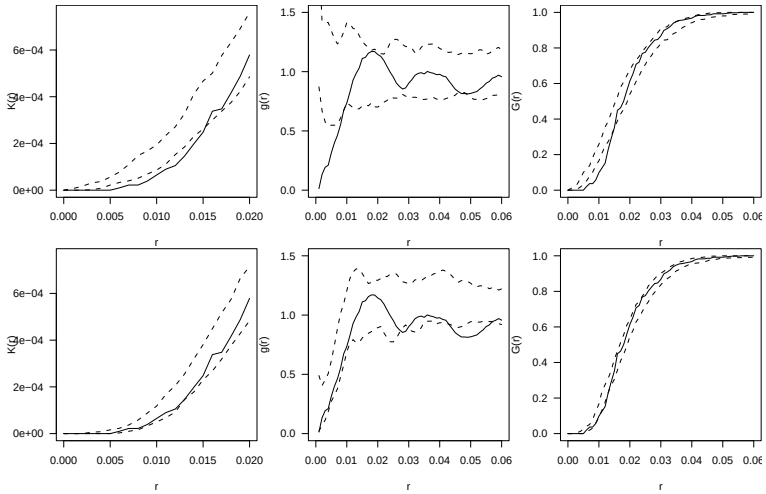

Figure 6.2: In both rows from left to right: plots of the inhomogeneous K, g summary statistics and the G summary statistic. Top row: data (solid line) together with envelopes (dashed lines) based on 39 simulation of an inhomogeneous Poisson process with intensity function as in right right plot of Figure 6.1. Bottom row: data (solid line) together with envelopes (dashed lines) based on 39 simulation of a pairwise interaction process estimated in the text. For more details on the summary statistics see, Baddeley et al. (2000) and Stoyan et al. (1995).

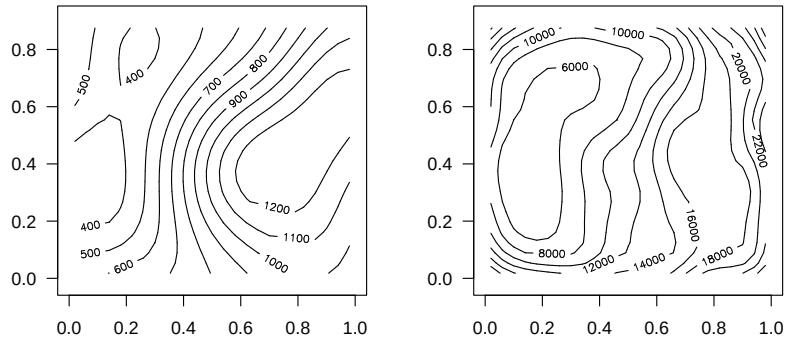

Figure 6.3: Top left: posterior mean of $\beta(\cdot, \psi)$. Top right: posterior variance.

linear relation $\omega(\theta) = 1 - R \times 0.11/0.015$ which is valid for $R \leq 0.13$. The latter restriction is no problem if we set $b_R \leq 0.13$. Furthermore, the interaction range R is expected to be much smaller than 0.13.

Using the sampling scheme described in Section 6.3 we have generated 250,000 updates skipping the initial $5,000$ updates as burn-in. Parameters for the sampler were $p = (0.35, 0.35, 0.1, 0.07, 0.07, 0.06)$, $\epsilon_R = 0.001$, $\epsilon_{R_0} = 0.0002$, $\epsilon_m = 0.5$, $\sigma_m = 0.01$, $b_{R_0} = 1.34 \times 10^{-3}$, $b_R = 0.05$, $b_{m_1} = 1$ and $b_{m_2} = 10$. The mean acceptance probabilities for the 6 updates are $0.46, 0.46, 0.48, 0.44, 0.46$ and 0.39, respectively.

Some results of sampling from the posterior are shown in Figure 6.3 and Figure 6.4. The posterior mean for $\beta(\cdot; \psi)$ shown in the left plot in Figure 6.3 is not unlike the estimated cell intensity in right plot of Figure 6.1 except that the posterior mean is higher. The variance for the prior intensity function is around 4×10^4. The posterior variance shown in Figure 6.3 is in the interval 5×10^3 to 3.5×10^4.

The empirical posterior distributions for $n(\psi)$, R, R_0 and m are shown in Figure 6.4. The corresponding posterior means are 193.8, 0.013, 0.0013 and 1.8. The posterior distributions of R, R_0 and m are quite different from the uniform prior distributions. This suggests that our prior has contributed little information compared to data.

The envelopes in the bottom row in Figure 6.2 are based on samples from a model with density (6.2) where the posterior means are used as the estimates of β, R, R_0, and m. This model is a clear improvement over the inhomogeneous Poisson process but still the summary statistics for the data are not quite inside the envelopes.

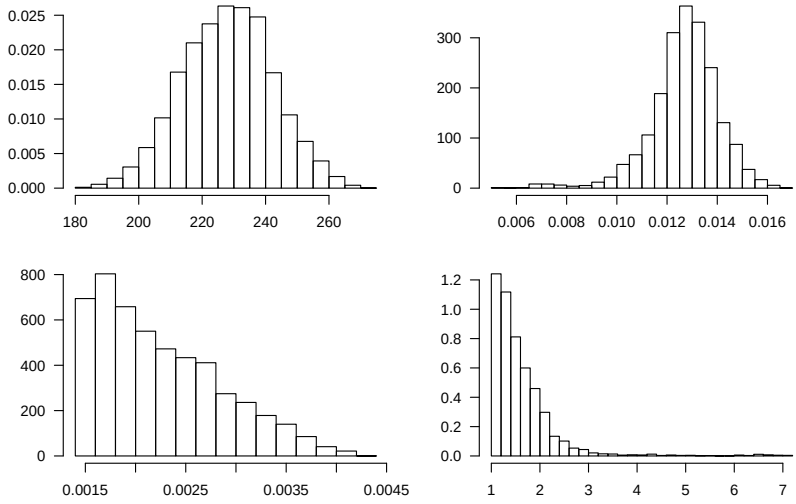

Figure 6.4: Empirical posterior distributions. Top left: number of kernels $n(\psi)$. Top right: interaction range R. Bottom left: hard core R_0. Bottom right: model parameter m.

Bibliography

Baddeley, A., Møller, J. & Waagepetersen, R. (2000). Non- and semi-parametric estimation of interaction in inhomogeneous point patterns, *Statist. Neerlandica* .

Berthelsen, K. K., Møller, J., Pettitt, A. N. & Reeves, R. (2003). Markov chain Monte Carlo without the normalising constant. Work in progress.

Cressie, N. & Davidson, J. L. (1998). Image analysis with partially ordered markov models, *Comput. Statist. Data Anal.* **29**: 1–26.

Davidson, J. L., Cressie, N. & Hua, X. (1999). Texture synthesis and pattern recognition for partially ordered markov models, *Pattern Recognition* **32**: 1475–1505.

Diggle, P. J., Fiksel, T., Grabarnik, P., Ogata, Y., Stoyan, D. & Tanemura, M. (1994). On parameter estimation for pairwise interaction point processes, *Int. Statist. Rev.* **62**: 99–117.

Gelman, A. & Meng, X.-L. (1998). Simulating normalizing constants: from importance sampling to bridge sampling to path sampling, *Statist. Sci.* **13**: 163–185.

Geyer, C. J. & Møller, J. (1994). Simulation procedures and likelihood inference for spatial point processes, *Scand. J. Statist.* **21**: 359–373.

Green, P. (1995). Reversible jump MCMC computation and Bayesian model determination, *Biometrika* **82**: 711–732.

Hahn, U., Jensen, E. B. V., van Lieshout, M.-C. & Nielsen, L. S. (2003). Inhomogeneous spatial point processes by location dependent scaling, *Adv. Appl. Prob.* **35**: 319–336.

Heikkinen, J. & Arjas, E. (1998). Non-parametric Bayesian estimation of a spatial Poisson intensity, *Scand. J. Statist.* **25**: 435–450.

Jensen, E. B. V. & Nielsen, L. S. (2000). Inhomogeneous Markov point processes by transformation, *Bernoulli* **6**: 761–782.

Kendall, W. S. & Møller, J. (2000). Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes, *Adv. Appl. Prob.* **32**: 844–865.

Møller, J. (2003). Shot noise Cox processes, *Adv. Appl. Prob.* **35**: 614–640.

Nielsen, L. S. (2000). Modelling the position of cell profiles allowing for both inhomogeneity and interaction, *Image Anal. Stereol.* **19**: 183–187.

Stoyan, D., Kendall, W. S. & Mecke, J. (1995). *Stochastic Geometry and Its Applications*, 2nd edn, Wiley, Chichester.